IN VIVO FUNCTIONS OF THE ETS FAMILY OF TRANSCRIPTION FACTORS: GENOMIC TARGETS OF ETS1

by

Katherine Jayne Chandler

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Oncological Sciences

The University of Utah

December 2012

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Katherine Jayne Chandler**

has been approved by the following supervisory committee members:

Barbara Graves	, Chair	**06/05/2012** Date Approved
Bradley Cairns	, Member	**06/05/2012** Date Approved
Dean Tantin	, Member	**06/05/2012** Date Approved
Matthew Topham	, Member	**06/05/2012** Date Approved
Katharine Ullman	, Member	**06/05/2012** Date Approved

and by **Bradley Cairns**, Chair of the Department of **Oncological Sciences**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

The regulation of gene expression is central to cell biology. While gene expression is modulated at many levels, the interaction between DNA and site-specific DNA binding transcription factors is a critical step. In metazoans, transcription factors usually exist in highly related families that exhibit a conserved preference for a particular DNA binding site. The ETS family provides a model for understanding how transcription factor families achieve distinct functions despite high conservation. With a focus on ETS1, genomic studies have identified two modes of ETS factor binding: redundant binding with other ETS factors, and specific binding of ETS1 only.

The work presented in this dissertation uses genomic data from massively parallel sequencing experiments to refine the model of ETS factor occupancy *in vivo*. ETS-redundant and ETS1-specific binding events were found to be correlated with distinct recruitment motifs. ETS-redundant binding is co-incident with histone marks associated with active promoters while ETS1-specific binding is co-incident with histone marks associated with distal transcriptional enhancers. ETS1 co-localizes with the transcriptional co-activator CBP at enhancers, but not at promoters, suggesting different mechanisms of ETS1 function at these two recruitment sequences.

Genome-wide disruption of ETS1 binding results in distinct expression profiles for genes near ETS-redundant and ETS1-specific binding events, confirming functional differences. At redundant promoters, disruption of ETS1 occupancy also provided the first example of dynamic time sharing by ETS1 and GABPA, establishing the molecular basis of redundancy within the ETS family. Unexpectedly, genes near ETS1-specific binding events were upregulated upon ETS1 disruption, suggesting a previously unidentified mode of regulation at these targets. These results serve as a framework for future studies investigating the specificity and redundancy of ETS factors *in vivo*.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ACKNOWLEDGMENTS

I would like to thank my advisor, Barbara Graves, for her super-human patience, dedication, and fantastic mentorship. Barbara has the ability to inspire her students to persevere in light of the challenges of bench science. I would like to thank members of my committee, Brad Cairns, Matt Topham, Dean Tantin, and Katie Ullman for their valuable knowledge and expertise to the furthering of this work, particularly in its later stages. I would like to thank past Graves Lab members Mary Nelson, Adam Blazjack, and Bargavi Thyagarajan for their scientific enthusiasm and support. I'd like to particularly thank past lab member Pete Hollenhorst for his technical and intellectual guidance through the first several years of my graduate work.

I would like to thank all of the wonderful friends that I have made in the lab, particularly Jahnvi Purwar, Ravi Alla, and Christian Pfluger, who kept me entertained and well caffeinated in the long stretches of graduate school. I would like to thank all of the members of my family, particularly my parents, who never stopped asking me when I'd get a job. Most importantly, I'd like to thank my boyfriend Jeff Sherlock, who supported me endlessly during the completion of this work, and has kept my life interesting with amazing travels and adventures.

CHAPTER 1

INTRODUCTION

Transcriptional regulation and transcription factors

A key step in gene expression is the interaction between a DNA binding transcription factor and a specific recognition sequence. This genome-encoded recognition process facilitates the recruitment and assembly of basal transcriptional machinery, which includes RNA polymerase and a variety of co-factors, including co-activators, co-repressors, and chromatin remodeling complexes, which ultimately regulate the transcription of RNA from the genomic template. Within the context of a cell, gene transcription must be highly regulated to respond to differing needs throughout the cell cycle. Within the context of an organism, additional layers of control must exist to establish and maintain a wide range of cell types. Gene expression is modulated at many points in this biological regulation, but transcription factor binding is a crucial and often misregulated step, making an understanding of transcription factor function central to many aspects of cell biology and pathogenesis.

This thesis focuses on the transcriptional regulation of mRNA by RNA polymerase II (RNAPII) in the context of human cells (Figure 1.1). Using the ETS family of transcription factors as a model, this work considers the ways in which divergent DNA binding sites, uniquely interacting protein-partners, and post-translational modifications guide specific genomic occupancy and regulatory function of transcription factors *in vivo*.

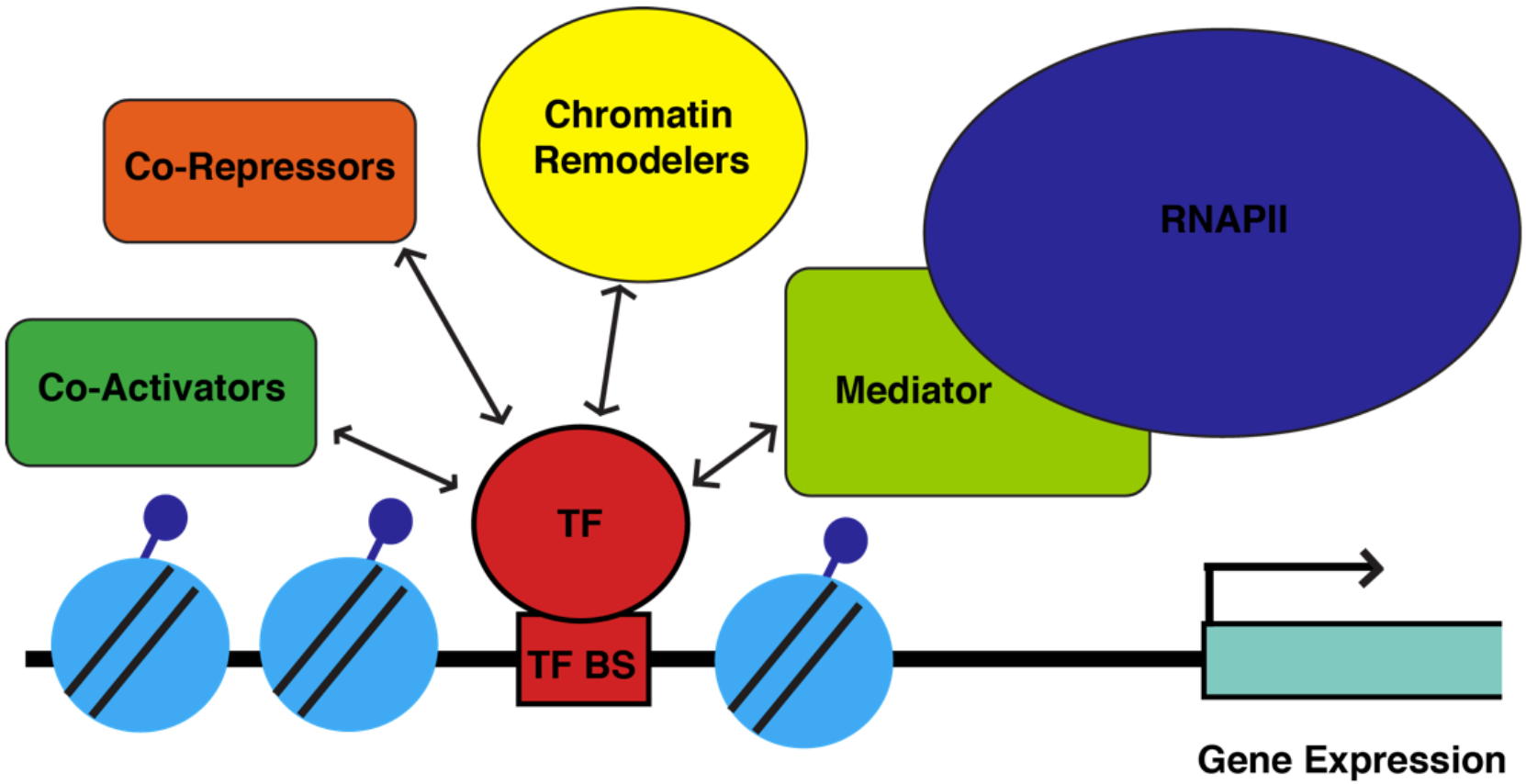


Figure 1.1 Eukaryotic transcription DNA binding transcription factors (TFs) bind to complimentary DNA binding sites (BS) in the genome. TFs can mediate the recruitment of a variety of transcriptional regulators, including co-activators, co-repressors, chromatin remodelers, and mediator proteins. These factors, as well as epigenetic marks like as histone modifications and DNA methylation, can facilitate the recruitment basal transcriptional machinery and influence gene transcription catalyzed by RNAPII.

The ETS family of transcription factors

As higher organisms have evolved, many, if not all, DNA binding transcription factors have undergone gene duplication events. As a result, the genomes of many organisms contain multiple members of a transcription factor family that are characterized by various degrees of conservation. Functional domains, such as DNA binding domains, are subject to selective conservation during evolution, thus many transcription factor families exhibit a conserved preference for DNA binding sequences [1]. In spite of having a shared binding site preference, there is evidence that within many transcription factor families individual members mediate distinct functions [2,3,4,5,6]. The work presented in this thesis explores the ways in which family members with highly conserved DNA binding domains identify and bind to distinct DNA motifs in the genome to perform unique biological functions.

One such family of regulatory proteins is the ETS family of transcription factors (Figure 1.2). The ETS family is characterized by a highly conserved DNA binding domain and is found in all metazoans; there are 8 *ets* genes in *Drosophila melanogaster* [7], 10 in *Caenorhabditis elegans* [8], and 28 in *Homo sapiens* [6]. Genetic studies in mouse models show distinct defects in mice with an individual ETS factor knocked out. Mouse phenotypes range from male infertility upon ETV4/5 knockout [9,10] to T cell and NK cell defects upon ETS1 knockout [11,12]. Many of these mouse genetic experiments have been recapitulated in flies and worms, and suggest that individual members of the ETS

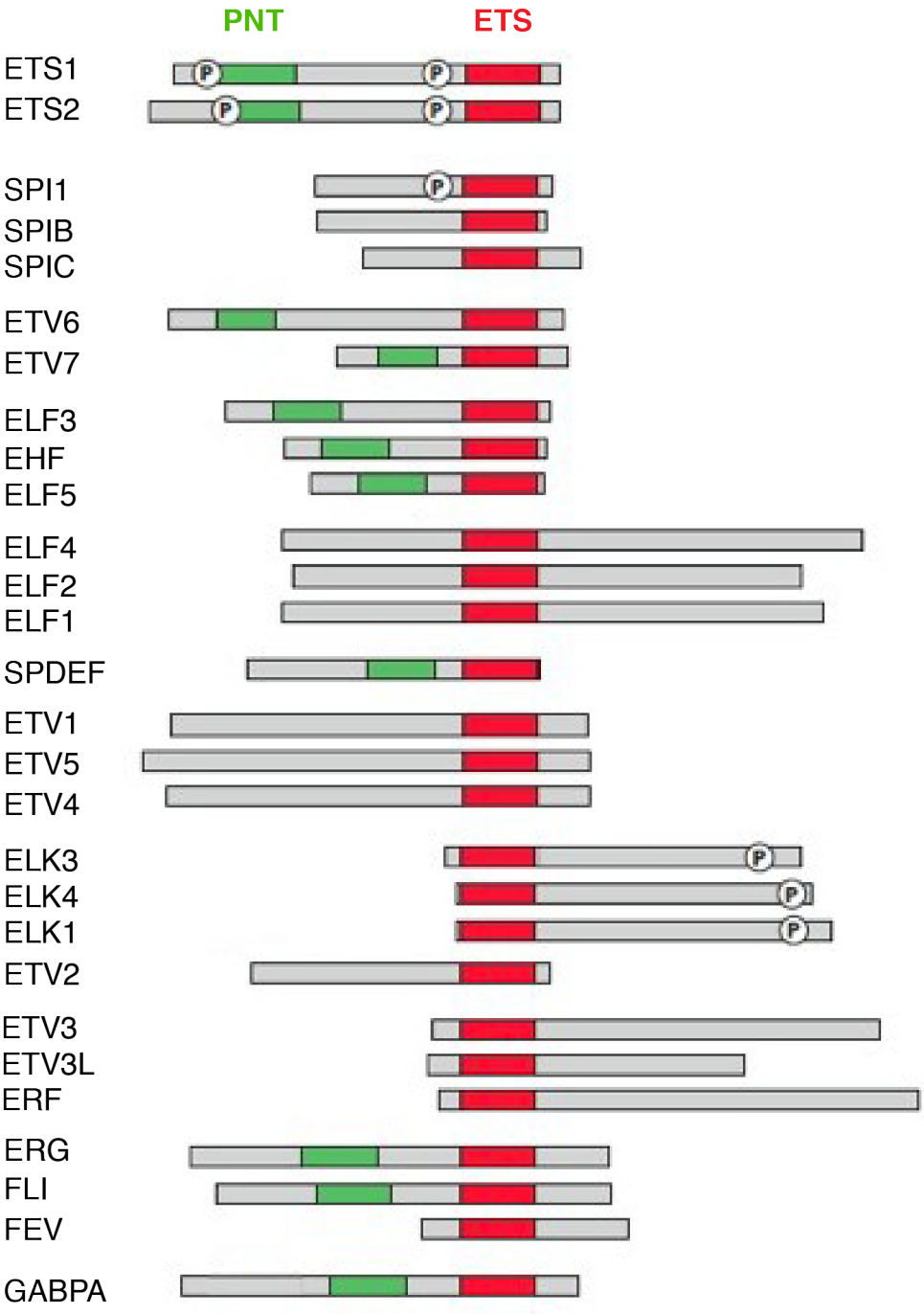


Figure 1.2 The ETS family of transcription factors The ETS family of DNA binding transcription factors contains a highly conserved DNA binding domain (the ETS domain, red), which binds to a core 5'-GGAA/T-3' DNA motif. Nearly 1/3 of ETS factors (11/28) also contain a conserved structural domain termed the Pointed Domain (the PNT domain, green), which mediates protein-protein interactions. Several ETS factors are phosphorylated in response to signaling, as denote by a circled P. (Figure adapted from Hollenhorst et al. [6])

family mediate distinct biological activity and regulate at least some unique genomic targets.

The ETS family specificity conundrum

Genetic experiments suggest discreet functions for ETS factors, yet multiple ETS family members are expressed within a single cell type [13,14]. This seeming incongruity frames a central conundrum in transcriptional regulation. How do closely related factors, co-expressed within a single cell, select unique transcriptional targets and maintain specificity of function?

The simplest explanation for specificity of function is that ETS family members have preferences for distinct DNA motifs. Over the past two decades many groups have performed site selection experiments to determine the preferred DNA binding sites of ETS family members [15,16,17,18,19,20,21,22]. Through systematic evolution of ligands by exponential enrichment (SELEX) and other multiplex binding assays, these studies have identified a highly conserved ETS binding site, 5'-GGAA/T-3, with variable but largely conserved flanking base that provide cooperative increases in DNA binding affinity [22]. While subtle differences have been identified between family member binding sites, it is more notable how strikingly similar the sequence preferences are across the family.

A recent study using protein binding microarrays to identify the DNA binding site preferences of 27 human ETS factors identified subtle but distinct site preferences within the four clades of the ETS family [23] (Figure 1.2). For example, ETS factors in Clade III prefer a 5'-GGAA-3' binding site while those in

Clade IV prefer a 5'-GGAT-3'. However, the differences in preferred binding sites were subtle and within a given clade there are often diverse functions. For example, Clade I contains the PEA3, TCF, ETS, ERF, and ERG subfamilies, representing over half of all known ETS factors and a diverse set of biological functions.

It is clear that DNA binding sites play a role in determining specific ETS factor function, but *in vitro* identification of high-affinity sites is not sufficient to characterize this mechanism. The above studies do not address binding preference for low- and medium-affinity binding sites, and frame the possibility that unique target selection by ETS factors may be mediated by sequences that diverge from the strongest, consensus sites. To address this possibility, ETS factors must be considered in the biological context of their structural domains, protein partners, and post-translational modifications, and ultimately occupancy should be determined *in vivo*.

Structured domains within the ETS Family

Many structured domains have been identified in the ETS family, but three structural domains are most common in the family: the ETS domain, the DNA binding domain that defines the family; the PNT domain, an interaction domain possessed by ~1/3 of the family; and autoinhibitory domains, which while not necessarily structurally conserved, represent an interesting regulatory feature in the family.

The ETS domain

All ETS factors contain a conserved DNA binding domain, the ETS domain, consisting of ~85 amino acids that fold into a winged-helix-turn-helix structure [6,24]. The ETS domain selectively binds to DNA motifs based on discrimination of 9 base pairs, and 5'-ACCGGAAGT-3' is the most commonly preferred site [23].

This domain is largely invariant, though there are examples of protein partners interacting specifically with the ETS domain of a particular factor, such as the interaction of PAX5 with ETS1 on the *mb-1* promoter [25,26]. Within an ETS factor, *cis*-interactions can also influence ETS domain binding to DNA. The ETS factors ELK1 and ELK4 contain identical DNA binding domains, but non-conserved residues distal to the ETS domain distinct conformations of the factors upon association with the consensus DNA binding site [27,28]. While this study considered ELK1 and ELK4 binding on the same site, it does highlight the possibility that *cis*-interactions with the ETS domain could mediate distinct binding preferences.

The PNT domain

In addition to the characteristic DNA binding domain, 11 of the 28 human ETS proteins contain a domain called the Pointed domain, or the PNT domain, named after the ETS1 *Drosophila* ortholog, Pointed-P2 [29]. The PNT domain term is used exclusively for ETS factors, the fold [30,31,32] is structurally similar to that found in the sterile α-motif domain, or SAM domain [33,34,35]. SAM

domains have been shown to interact with SAM and non-SAM protein domains, as well as with RNA and other molecules [36], linking the ETS family evolutionarily to proteins of diverse function.

The PNT domains of multiple ETS factors interact with the homologous lysine acetyl-transferases (KATs) CBP/p300 [37,38,39] [40] [41]. The interaction between ETS-PNT and CBP/p300 is essential for the transcriptional function of many ETS factors, presumably because histone acetylation establishes a transcriptionally permissive environment and mediates the recruitment of the basal transcriptional machinery. It was observed that ETS1-driven transcription is enhanced by RAS- mitogen-activated protein kinase (RAS-MAPK) signaling [42], and later shown that ETS1-PNT serves as a kinase docking site for ERK1/2 [43]. Upon docking at ETS1-PNT, ERK1/2 can phosphorylate ETS1 at two residues N-terminal to the PNT-domain [44,45]. This phosphorylation enhances the interaction of ETS1-PNT with the homologous lysine acetyl-transferases (KATs) CBP/p300 [46], increasing the transactivation function of ETS1, further building on the significance of the ETS-PNT.

Other co-factors have been implicated in interactions with ETS1-PNT, such as the ubiquitin ligase UBC9 [47] and the histone deacetylase Daxx [48]. Additionally, the PNT of the ETS factor ETV6 is capable of homotypic polymerization [49], facilitating cooperative DNA binding and perhaps by overcoming the factor's relatively low DNA binding affinity [50].

Autoinhibitory domains in ETS factors

The term autoinhibition (AI) refers to the attenuation of DNA binding by a *cis*-element, and while AI domains are not necessarily conserved structures, they are widely used across the ETS family. ETS factors that regulate through AI include ETV4 [51,52], ETV5 [53], ETV6 [50], and members of TCF sub-family [54].

The best characterized case of AI in the ETS family, however, is ETS1 [55], and the AI domain is a dynamic point of regulation for the ETS1 protein. The AI domain of ETS1 is signaling responsive [56,57,58,59], is alternatively spliced in different isoforms of ETS1 [60,61,62], and is thought to interact with the transcription factor RUNX1 [63]. Taken together, the diversity of functional mechanisms derived from the ETS1 AI domain illustrate the number of permutations available to fine tune transcriptional function and drive DNA binding site specificity.

Protein partners of ETS1

The ETS family has been implicated in numerous protein partnerships, which can serve to target particular ETS factors to unique genomic targets. Some partnerships are shared among family members, such as the interaction of PU.1, ERG, ELF1, and other ETS factors with the transcription factor JUN [64,65,66]. Other partnerships are unique to a particular factor, such as the interaction of GABPA with its non-DNA binding subunit GABPB [67]. While many factors have been shown to interact with ETS1, here we focus here on two partners that have

been shown to be relevant in the context of DNA motif specificity: RUNX1 and PAX5.

ETS1 forms a protein partnership with RUNX1

RUNX1 is a member of the Runt-family of transcription factors, and is important for hematopoietic development and B cell function. I*n vitro* DNA binding and biochemical studies have shown that ETS1 and RUNX1 cooperatively bind to a motif found in the MoMLV enhancer [63]. This enhancer contains the motif 5'-CAGGATATCTGTGGTAAGC -3', which is comprised of a weak ETS motif (GGAT) and a degenerate RUNX motif (TGTGG). While the individual sites are unfavorable, *in vitro* and *in vivo* occupancy has been observed for ETS1 and RUNX1 at this motif. Truncation experiments show that cooperativity at this site requires the ETS1 AI domain, suggesting that RUNX1 relieves ETS1-autoinhibition, though a direct physical interaction between RUNX1 and the ETS1 AI domain has not been defined [63]. In addition to facilitating cooperative DNA binding by unphosphorylated ETS1, it has been reported that RUNX1 interacts with the phosphatase calcineurin, which de-phosphorylates ETS1-SRR [68], though it should be noted that these experiments involved overexpression of calcineurin at nonphysiological levels.

ETS1 forms a protein partnership with PAX5

Like RUNX1, PAX5 is a transcription factor important for B cell function. Through a cooperative interaction with the ETS domain [25], ETS1 and PAX5

bind to a highly unfavorable composite motif in the *mb-1* promoter to drive gene expression [25,26,69]. This interaction is mediated by four residues in the ETS1-ETS domain, and ETS factors containing three out of four of these conserved residues, including ELK1 and ELK3, were also shown to interact with PAX5 [69]. These interactions were less efficient, however, and the investigators postulate that the strength of the ETS1-PAX5 interaction encourages ETS1 specificity at the composite site *in vivo*. These two examples build the case that combinatorial control could provide a route to specificity, as partner proteins can facilitate ETS factor binding at composite sites that diverge from the consensus.

ETS1 responds to several signaling pathways

Cellular signaling regulates transcription factor function. By modulating nuclear localization, DNA binding affinity, and/or interactions with other transcriptional regulatory proteins, transcription factors can respond to diverse signals with a high degree of fine tuning [70]. Here we focus on RAS-MAPK and calcium signaling, both of which impinge on the ETS1 protein (Figure 1.3).

ETS1 responds to RAS-MAPK signaling

The RAS-MAPK pathway is a signaling cascade that impinges on multiple ETS family members to affect transcriptional function. Upon growth-factor stimulation of membrane bound receptors, the RAS GTPase transmits the extracellular signal from receptor tyrosine kinases (RTKs) through a kinase cascade to MAPKs. Activated nuclear MAPKs can phosphorylate target

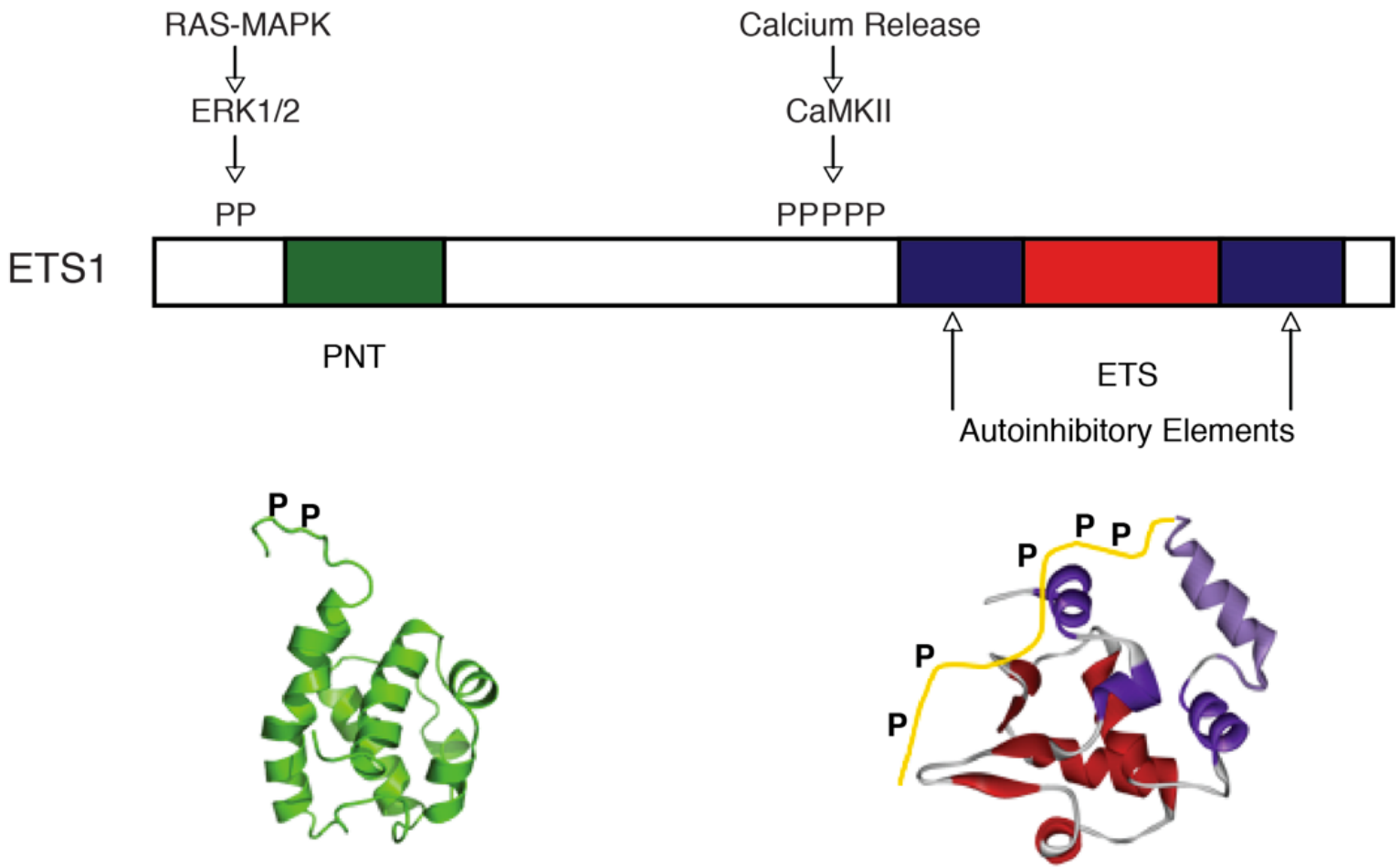


Figure 1.3 Signaling to ETS1 ETS1 is impinged on by multiple signaling path ways. RAS-MAPK signaling to ERK1/2 results in phosphorylation of two discreet residues near the ETS1-PNT domain [42-44]. This phosphorylation mark results in an enhanced interaction between ETS1 and CBP/p300 [46], which increases the transcriptional activity of ETS1. Calcium dependant signaling activates CaMKII, which phosphorylates ETS1 in a flexible serine rich (SRR) region near the ETS domain, resulting in a 10-fold decrease in DNA binding affinity [56-58].

transcription factor substrates, connecting the initial extracellular stimuli to a transcriptional readout.

Upon RAS-signaling, the MAPK ERK1/2 phosphorylates ETS1 and ETS2 in an unstructured region N-terminal to the PNT-domain. This phosphorylation event induces a conformational change in the PNT domain and enhances the interaction affinity between PNT and CBP/p300, thus affecting ETS1 driven transcription [42,44,45,46]. Similarly, MAPK phosphorylation of the ETS family member ELK1 enhances transcription through an interaction with CBP/p300 [71,72]. Upon ELK1 phosphorylation, the conformation of this interaction changes, increasing the DNA binding affinity of ELK1 and facilitating enhanced KAT activity of CBP/p300 [73]. These two examples of RAS signaling to ETS factors feature many of the same players, but utilize different mechanisms to respond to signaling, underscoring the use of signaling to functionally distinguish similar factors.

ETS1 responds to calcium signaling

Extracellular signals are transmitted to the nucleus is through regulation of internal stores of calcium as a signaling molecule. Upon receptor engagement, the enzyme phospholipaseC (PLC) is activated and catalyzes generation of the second messenger molecules diacylglycerol (DAG) and inositol trisphosphate (IP_3). IP_3 then binds to an endoplasmic reticulum (ER) bound IP_3–receptor and stimulates the opening of ER Ca^{2+} channels, resulting in a large flux of cytoplasmic calcium ions. The increase in cytoplasmic calcium activates the

calcium/calmodulin-dependent protein kinases II (CaMKII). Activated CaMKII can phosphorylate ETS1 at multiple residues in the SRR flanking the C-terminal ETS domain [56,58]. This phosphorylation stabilizes the AI domain, resulting in an up to 50-fold inhibition of DNA binding affinity [57,58,59]. The extent to which DNA binding affinity is decreased is directly correlated with the number of phosphates modifying the SRR, serving as a rheostat to fine tune ETS1 binding affinity [58]. Using artificial reporter constructs, calcium signaling has been shown to reduce the expression of the ETS1 target GS-CSF [74], but the comprehensive biological relevance of this signaling was not well understood.

Combining specificity mechanisms diversifies functionality

The above discussion considered mechanisms of achieving specificity as discrete functions, but in many cases these mechanisms are used in combination to further regulate transcriptional function. This provides the ETS family, and transcription factors in general, with a large arsenal of tools to achieve specificity. For example, ETS1 is regulated by autoinhibition; calcium signaling reinforces autoinhibition; and a protein partnership with RUNX1 relieves autoinhibition. Each of these three mechanisms represents a layer of specificity that can individually regulate ETS1 function, or can combinatorially fine tune ETS1 specificity.

In vivo occupancy of transcription factors

Thus far, we have considered *in vitro* biochemical studies to examine functionality of the ETS family. In order to understand this functionality in a

biological context, however, we must understand the genomic occupancy of ETS factors *in vivo*. Within a cell, what are the genes that a particular ETS factor regulates, and is this regulation exclusive for that ETS factor or redundant across the family?

Until recently, transcriptional regulatory elements were determined by using artificial reporter constructs in *in vitro* assays. By positioning a putative regulatory region upstream of a reporter gene and making mutations or truncations, the *cis*-elements driving transcription and in many cases the associated DNA binding protein, could be identified. While these techniques illuminated many of the basic principles of transcriptional regulation, they are time consuming and labor intensive. Identification of regulatory elements was limited to a small subset of genes, and consideration was limited to 1-2 kilobases (kb) flanking the transcription start site (TSS). Further, these studies relied on exogenous reporter constructs that do not necessarily recapitulate endogenous loci in terms of chromatin state, genomic position, and interactions with distal regulatory regions.

While a handful of ETS-target genes were identified using basic genetic approaches, the vast majority of ETS factor targets remained unknown until the development of the chromatin immunoprecipitation (ChIP) technique, which maps the *in vivo* association of endogenous factors with DNA [75]. The technique involves chemically crosslinking live cells to stabilize chromatin, shearing crosslinked chromatin to 300-1000 base pair fragments immunoprecipitating with

an antibody against the factor of interest, and reversing crosslinks to isolate the associated DNA for PCR or other analysis. By combining ChIP with microarray technology (ChIP-on-chip) [76] or more recently with massively parallel sequencing techniques (ChIPSeq) [77], transcription factor occupancy can be identified *in vivo* on a large scale. An important distinction between ChIP-on-chip and ChIPSeq is that due to the size of the human genome, the former requires tiling of a discrete set of known genomic elements on a microarray, usually regions -5kb to +2kb around a TSS, while the latter has the advantage of *de novo* sequencing and can thereby identify previously unknown regulatory elements.

Genomic targets of the ETS Family

To determine the transcriptional targets of ETS1 and other ETS factors, ChIP-on-chip using promoter microarrays was performed for the ETS factors ETS1, GABPA, ELF1, and ELK1 [78,79]. Because ETS1 is predicted to have a T cell specific function [11,12], experiments were carried out in a human T cell line. Unexpectedly, a large subset of the identified ETS targets were not unique to a particular factor, and at many promoter regions redundant occupancy of all examined ETS factors was observed. Motif analysis revealed that redundantly occupied promoters contain a high-affinity consensus site, 5'-CCGGAA-3', and are frequently associated with the promoters of housekeeping genes. An ETS1 specific class was also identified in this study, and many of these regions were found to contain a lower affinity ETS site, often juxtaposed with a weak RUNX binding site [78]. These findings have been incorporated into the working model

of ETS factor function: high-affinity sites in the promoters of housekeeping (HK) genes mediate redundant ETS factor function and low-affinity composite sites mediate specific ETS factor functionality (Figure 1.4).

While this study revealed a great deal about the genomic occupancy of ETS factors, it opened up several broad questions. First, while a subset of ETS1 specific targets were identified, the neighboring genes were unrelated to the observed phenotype of the ETS1 knockout mouse, and it was unclear how – or if – ETS1 directly regulates the expression of T cell genes.

In addition to this question of specific ETS1 function, the surprising prevalence of redundant ETS factor occupancy raised a number of questions regarding the nature of redundancy. For example, do ETS factors time share an individual binding site in a single cell, or does redundancy reflect the average occupancy of that loci across all cells assayed? Further, is the function of ETS factors at redundant sites truly redundant, meaning that one ETS factor can fully compensate for the function of another ETS factor at these sites?
Reporter constructs have shown transcriptional functionality of redundant ETS binding events [80], but these assays lack endogenous context. Correlating endogenous gene expression with *in vivo* factor occupancy provides a biologically relevant way to confirm the functionality of DNA binding events. For example, gene expression studies have been used to identify previously unknown functional mechanisms for estrogen receptor (ER) mediated down regulation [81,82] and have led to the validation of numerous p53

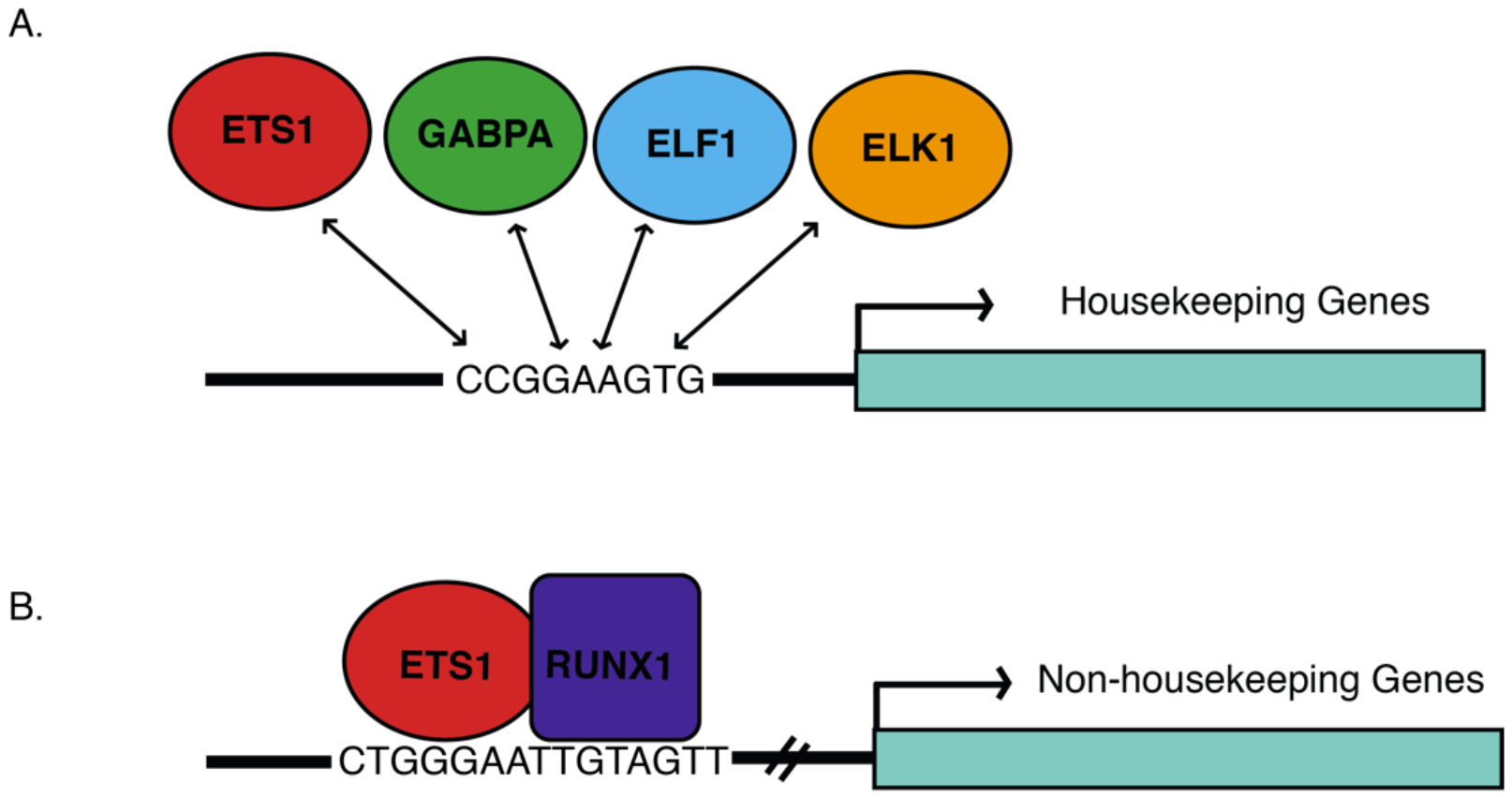


Figure 1.4 Genomic Targets of the ETS family ETS factors are recruited to DNA binding sites *in vivo* through two distinct mechanisms: redundant recruitment, which does not discriminate between ETS factors, and specific recruitment, which is selective for an individual ETS factor. (A) Redundant ETS factor recruitment is mediated by a high-affinity binding site that strongly resembles the *in vitro* derived ETS consensus site [78, 79]. This type of binding event frequently occurs < 25 bp from the TSS of a HK gene. (B) Specific ETS factor recruitment is mediated by a lower-affinity binding site, and usually occurs distal to neighboring gene promoters. For ETS1, specific recruitment is often observed at an ETS site juxtaposed with a RUNX binding site [78].

transcriptional targets that *in vitro* assays could not have identified [83]. However, because of inherent inconsistencies in probe quality, microarray experiments require comparison to a reference state, making it difficult to quantify the expression of genes which are expressed across nearly all cell types and in most cellular conditions. Thus, the *in vivo* functionality of redundant ETS binding events is unclear.

Transcription factors may bind several distinct sites

Different transcription factor families have different requirements for recruitment motifs. The occupied loci of some factors, such as the STAT [77] and ETS families [78], are enriched for sequences similar to known binding sites previously identified *in vitro*. Other factors, such as the E2F family, show almost no proclivity for the high affinity sequences derived *in vitro* or any other site [84], leaving open the question of how E2F factors select their targets *in vivo*.

Of factors that do preferentially bind a known consensus sequence, there is no doubt diversity in the sites occupied. An experiment performed using protein-binding microarrays and 104 mouse transcription factors found that half of the queried factors bound two or more distinctly different sites with high-affinity [85]. Crystallographic experiments have shown that the hormone responsive transcription factor glucocordicoid receptor (GR) adopts different structural conformations based on binding site differences as subtle as a single nucleotide [86], and that these conformational differences impact GR transcriptional function. This example highlights that dual- or multi- site recognition by a

transcription factor could provide an additional control for functional regulation of transcription, and *in vivo* determination of factor occupancy is likely to reveal a large complement of sites used by transcription factors.

Genome organization

Another insight revealed by genomic experiments and bioinformatics is that very few of the possible genomic binding sites for a particular factor are occupied *in vivo* [76,87,88,89]. While the inherent affinity of a transcription factor for a binding site certainly plays a role in determining occupancy, genome organization is equally important. Genetic material is highly packaged, with genomic DNA wrapped around histone octamers, or nucleosomes, to form chromatin [90,91]. The structure of this packaging can be modulated to make particular regions of the genome more or less accessible and dictates where, and with what affinity, transcription factors bind DNA. To a large extent, differences in the accessibility of the genome are what lead to the diversity of gene expression profiles across different cell types [92]. There are a number of factors that can influence the relative accessibility of DNA, including histone marks and DNA methylation. The interplay of these factors can establish a wide range of chromatin environments, guiding transcription factor occupancy and ultimately gene expression.

Gene regulatory elements

It has long been known that the chromatin packaging of gene promoters, or regions proximal to the TSS of RNAPII transcribed genes, is generally open and is therefore accessible to transcription factors. Because these regions are accessible, they are susceptible to DNase cleavage assays, and DNaseI-hypersensitivity (DNaseHS) is often used as a readout to identify novel genes and regulatory elements [93,94]. Though proximal promoters are the most well studied type of regulatory element, genome-wide DNaseHS studies have highlighted the abundance of other types of functional elements [95]. Transcriptional enhancers, distal elements that promote transcription, are also regions of open chromatin and are more prevalent in the genome than promoters. DNaseHS assays in multiple cell types reveal that while promoter regions are near- universally open, the accessibility of enhancer regions varies by cell type [96], suggesting that distal enhancers, not promoters, define cell type specific expression programs. As discussed above, ETS factor occupancy had been determined at transcriptional promoters as well as at enhancers, but it was not clear if ETS factors play the same role in these two different contexts.

Histone marks

Histone tails can accept many different covalent modifications, including acetylation, methylation, phosphorylation, ubiquitylation, and sumoylation. Promoter and enhancer elements are correlated with a specific signature of histone marks that regulate local accessibility. Histone marks can regulate

chromatin environment directly by disrupting DNA-nucleosome contacts or indirectly by recruiting chromatin modifying factors.

Histone acetylation is invariably associated with transcriptional activation, facilitated through the 'puffing' or de-condensing of acetyl-rich chromatin. Histone acetyltransferases (HATs) are grouped into three main families, GNAT, MYST, and CBP/p300, and are generally promiscuous in their targets (reviewed in [97]). The transcriptional effects of histone methylation are more diverse, with H3K4-, H3K36-, and H3K79-methylation being implicated in transcriptional activation and H3K9-, H3K27-, and H4K20-methylation being implicated in transcriptional repression (reviewed in [98]). Histone methyltransferases (HMTs) are typically substrate specific, and it is not unusual for a single HMT to deposit a single specific methyl mark.

Relative to the rest of the genome, promoters are depleted of nucleosomes, with a marked nucleosome free region at the TSS. The nucleosomes nearby are hypoacetylated and marked with H3K4me3 [99,100], a hallmark of active transcription. This mark is added by the HMT SET1 [101], which physically associates with active RNAPII [102]. Enhancers, on the other hand, are not marked with H3K4me3, but with H3K4me1 [103], highlighting the diversity and specificity of methyl-marks. Enhancers are also enriched for H3K9ac or H3K14ac [104], perhaps because of high CBP/p300 HAT occupancy, but these marks are not specific to enhancers and generally mark regions of active chromatin. The relationship, if any, between ETS factors and histone

marks had not been determined, and we postulated that correlation with unique histone marks could provide functional insight for different types of ETS factor targets.

DNA methylation

Like histones, DNA can be covalently modified as a mechanism of regulating transcription. DNA methyl transferases, DNMTs, can facilitate the addition of a methyl-group to the cytosine of 5'-CpG-3' dinucleotides, and this modification can repress transcription by recruiting additional repressive proteins or by disrupting transcription factor binding sites. This method of regulation is of particular interest in the context of the ETS family, because the ETS consensus site, 5'-CCGGAA-3', contains a CpG dinucleotide. Several ETS family members have been shown to be sensitive to site methylation [105,106](S.C. unpublished data). Based on this sensitivity DNA methylation could be a key factor in determining ETS factor occupancy at consensus sites. As mentioned above, the 5'-CCGGAA-3' ETS consensus site most frequently occurs in TSS proximal promoters, which are largely hypomethylated regions (reviewed in [107]). Though the causal relationship between ETS factor occupancy and hypomethylation has not been explored, the correlation is striking, and sets the stage for future examining the potential function of ETS factors at housekeeping promoters. One possibility is that ETS factors bind with high-affinity to CpG-containing sites to protect the site from methylation and thereby ensure constitutive expression, a mechanism which has been described for the transcription factor MLL, which

occupies a CpG-containing binding site in the Hoxa9 promoter to protect the site from methylation [108]. Because all ETS factors have the ability to bind high-affinity CpG-containing sites, we postulate that CpG protection could be the redundant function of ETS factors.

The focus of this thesis

The ETS family provides a model for understanding how transcription factor families achieve distinct functions despite high conservation. At the commencement of this work, analysis of ETS factor binding to promoter regions had identified two modes of ETS1 binding: redundant binding with other ETS factors, and specific binding of ETS1 only. The work presented in this dissertation uses data sets generated by massively parallel DNA sequencing techniques to refine the model of ETS factor occupancy *in vivo*, and highlights the structural, biological, and behavioral differences between different types of ETS1 occupied regulatory elements.

In Chapter 2, we established that ETS-redundant and ETS1-specific regulatory elements are associated with distinct chromatin landscapes, DNA recruitment motifs, and ontological gene sets. Using a genome wide approach, we identified a large number of previously unidentified transcriptional enhancers, which were found to be mediated by a degenerate ETS binding motif and often occur near T cell-specific genes. Overlaying histone mark profiles on ETS-factor occupancy data, we found that ETS-redundant and ETS1-specific regulatory elements are correlated with marks of active promoters and transcriptional

enhancers, respectively. We also found a striking co-localization of ETS1 and the transcriptional co-activator CBP at enhancers, but not at promoters, suggesting that ETS1 may behave differently at these two recruitment sequences.

In Chapter 3, we disrupted ETS1 binding genome wide to reveal distinct behaviors for ETS-redundant and ETS1-specific target genes and illustrating the separate biological functions of redundant and specific target genes. At redundant promoters, disruption of ETS1 occupancy has provided the first example of dynamic time sharing by ETS1 and GABPA, establishing the molecular basis of redundancy within the ETS family. Unexpectedly, expression analysis over a time course of ETS1 disruption revealed upregulation of genes near ETS1 specific binding events, suggesting a previously unidentified mode of regulation at these targets. These results have important implications for understanding how ETS1 functionally mediates different transcriptional programs, and suggests a model of ETS family function wherein specificity is a continuum rather than a dichotomy.

References

1. Nowick K, Stubbs L (2010) Lineage-specific transcription factors and the evolution of gene regulatory networks. Brief Funct Genomics 9: 65-78.

2. Guth SI, Wegner M (2008) Having it both ways: sox protein function between conservation and innovation. Cell Mol Life Sci 65: 3000-3018.

3. Suske G (1999) The Sp-family of transcription factors. Gene 238: 291-300.

4. Owen GI, Zelent A (2000) Origins and evolutionary diversification of the nuclear receptor superfamily. Cell Mol Life Sci 57: 809-827.

5. Rohs R, Jin X, West SM, Joshi R, Honig B, et al. (2010) Origins of specificity in protein-DNA recognition. Annu Rev Biochem 79: 233-269.

6. Hollenhorst PC, McIntosh LP, Graves BJ (2011) Genomic and biochemical insights into the specificity of ETS transcription factors. Annu Rev Biochem 80: 437-471.

7. Hsu T, Schulz RA (2000) Sequence and functional properties of Ets genes in the model organism Drosophila. Oncogene 19: 6409-6416.

8. Hart AH, Reventar R, Bernstein A (2000) Genetic analysis of ETS genes in C. elegans. Oncogene 19: 6400-6408.

9. Chen C, Ouyang W, Grigura V, Zhou Q, Carnes K, et al. (2005) ERM is required for transcriptional control of the spermatogonial stem cell niche. Nature 436: 1030-1034.

10. Laing MA, Coonrod S, Hinton BT, Downie JW, Tozer R, et al. (2000) Male sexual dysfunction in mice bearing targeted mutant alleles of the PEA3 ets gene. Mol Cell Biol 20: 9337-9345.

11. Muthusamy N, Barton K, Leiden JM (1995) Defective activation and survival of T cells lacking the Ets-1 transcription factor. Nature 377: 639-642.

12. Barton K, Muthusamy N, Fischer C, Ting CN, Walunas TL, et al. (1998) The Ets-1 transcription factor is required for the development of natural killer cells in mice. Immunity 9: 555-563.

13. Hollenhorst PC, Jones DA, Graves BJ (2004) Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. Nucleic Acids Res 32: 5693-5702.

14. Galang CK, Muller WJ, Foos G, Oshima RG, Hauser CA (2004) Changes in the expression of many Ets family transcription factors and of potential target genes in normal mammary tissue and tumors. J Biol Chem 279: 11281-11292.

15. Brown LA, Yang SH, Hair A, Galanis A, Sharrocks AD (1999) Molecular characterization of a zebrafish TCF ETS-domain transcription factor. Oncogene 18: 7985-7993.

16. John S, Marais R, Child R, Light Y, Leonard WJ (1996) Importance of low affinity Elf-1 sites in the regulation of lymphoid-specific inducible gene expression. J Exp Med 183: 743-750.

17. Poon GMK, Macgregor RB (2003) Base coupling in sequence-specific site recognition by the ETS domain of murine PU.1. J Mol Biol 328: 805-819.

18. Mao X, Miesfeldt S, Yang H, Leiden JM, Thompson CB (1994) The FLI-1 and chimeric EWS-FLI-1 oncoproteins display similar DNA binding specificities. J Biol Chem 269: 18216-18222.

19. Ray-Gallet D, Mao C, Tavitian A, Moreau-Gachelin F (1995) DNA binding specificities of Spi-1/PU.1 and Spi-B transcription factors and identification of a Spi-1/Spi-B binding site in the c-fes/c-fps promoter. Oncogene 11: 303-313.

20. Shore P, Sharrocks AD (1995) The ETS domain transcription factors Elk-1 and SAP-1 exhibit differential DNA binding specificities. Nucleic Acids Res 23: 4698-4706.

21. Nye JA, Petersen JM, Gunther CV, Jonsen MD, Graves BJ (1992) Interaction of murine ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. Genes Dev 6: 975-990.

22. Szymczyna BR, Arrowsmith CH (2000) DNA binding specificity studies of four ETS proteins support an indirect read-out mechanism of protein-DNA recognition. J Biol Chem 275: 28363-28370.

23. Wei GH, Badis G, Berger MF, Kivioja T, Palin K, et al. (2010) Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. EMBO J 29: 2147-2160.

24. Donaldson LW, Petersen JM, Graves BJ, McIntosh LP (1996) Solution structure of the ETS domain from murine Ets-1: a winged helix-turn-helix DNA binding motif. EMBO J 15: 125-134.

25. Wheat W, Fitzsimmons D, Lennox H, Krautkramer SR, Gentile LN, et al. (1999) The highly conserved beta-hairpin of the paired DNA-binding domain is required for assembly of Pax-Ets ternary complexes. Mol Cell Biol 19: 2231-2241.

26. Fitzsimmons D, Lukin K, Lutz R, Garvie CW, Wolberger C, et al. (2009) Highly cooperative recruitment of Ets-1 and release of autoinhibition by Pax5. J Mol Biol 392: 452-464.

27. Mo Y, Vaessen B, Johnston K, Marmorstein R (1998) Structures of SAP-1 bound to DNA targets from the E74 and c-fos promoters: insights into DNA sequence discrimination by ETS proteins. Mol Cell 2: 201-212.

28. Mo Y, Vaessen B, Johnston K, Marmorstein R (2000) Structure of the Elk-1-DNA complex reveals how DNA-distal residues affect ETS domain recognition of DNA. Nat Struct Biol 7: 292-297.

29. Klambt C (1993) The Drosophila gene pointed encodes two ETS-like proteins which are involved in the development of the midline glial cells. Development 117: 163-176.

30. Slupsky CM, Gentile LN, Donaldson LW, Mackereth CD, Seidel JJ, et al. (1998) Structure of the Ets-1 pointed domain and mitogen-activated protein kinase phosphorylation site. Proc Natl Acad Sci U S A 95: 12129-12134.

31. Tran HH, Kim CA, Faham S, Siddall M-C, Bowie JU (2002) Native interface of the SAM domain polymer of TEL. BMC Struct Biol 2: 5.

32. Kim CA, Bowie JU (2003) SAM domains: uniform structure, diversity of function. Trends Biochem Sci 28: 625-628.

33. Scholz H, Deatrick J, Klaes A, Klämbt C (1993) Genetic dissection of *pointed*, a Drosophila gene encoding two ETS-related proteins. Genetics 135: 455-468.

34. Thanos CD, Goodwill KE, Bowie JU (1999) Oligomeric structure of the human EphB2 receptor SAM domain. Science 283: 833-836.

35. Thanos CD, Faham S, Goodwill KE, Cascio D, Phillips M, et al. (1999) Monomeric structure of the human EphB2 sterile alpha motif domain. J Biol Chem 274: 37301-37306.

36. Qiao F, Bowie JU (2005) The many faces of SAM. Sci STKE 286: 10.1126/stke.2862005re7.

37. Yang SH, Yates PR, Whitmarsh AJ, Davis RJ, Sharrocks AD (1998) The Elk-1 ETS-domain transcription factor contains a mitogen-activated protein kinase targeting motif. Mol Cell Biol 18: 710-720.

38. Jayaraman G, Srinivas R, Duggan C, Ferreira E, Swaminathan S, et al. (1999) p300/cAMP-responsive element-binding protein interactions with ets-1 and ets-2 in the transcriptional activation of the human stromelysin promoter. J Biol Chem 274: 17342-17352.

39. Yamamoto H, Oikawa T (1999) Interaction of transcription factor PU. 1 with coactivator CBP. Tanpakushitsu Kakusan Koso 44: 1389-1395.

40. Kang HS, Nelson ML, Mackereth CD, Scharpf M, Graves BJ, et al. (2008) Identification and structural characterization of a CBP/p300-binding domain from the ETS family transcription factor GABP alpha. J Mol Biol 377: 636-646.

41. Janknecht R, Ernst WH, Pingoud V, Nordheim A (1993) Activation of ternary complex factor Elk-1 by MAP kinases. Embo J 12: 5097-5104.

42. Yang B-S, Hauser CA, Henkel G, Colman MS, Van Beveren C, et al. (1996) Ras-mediated phosphorylation of a conserved threonine residue enhances the transactivation activities of c-Ets-1 and c-Ets-2. Mol Cell Biol 16: 538-547.

43. Seidel JJ, Graves BJ (2002) An ERK2 docking site in the Pointed domain distinguishes a subset of ETS transcription factors. Genes Dev 16: 127-137.

44. Nelson ML, Kang HS, Lee GM, Blaszczak AG, Lau DK, et al. (2010) Ras signaling requires dynamic properties of Ets1 for phosphorylation-enhanced binding to coactivator CBP. Proc Natl Acad Sci U S A 107: 10026-10031.

45. McIntosh LP, Kang HS, Okon M, Nelson ML, Graves BJ, et al. (2009) Detection and assignment of phosphoserine and phosphothreonine residues by (13)C- (31)P spin-echo difference NMR spectroscopy. J Biomol NMR 43: 31-37.

46. Foulds CE, Nelson ML, Blaszczak AG, Graves BJ (2004) Ras/mitogen-activated protein kinase signaling activates Ets-1 and Ets-2 by CBP/p300 recruitment. Mol Cell Biol 24: 10954-10964.

47. Hahn SL, Wasylyk B, Criqui-Filipe P, Criqui P (1997) Modulation of ETS-1 transcriptional activity by huUBC9, a ubiquitin- conjugating enzyme. Oncogene 15: 1489-1495.

48. Li RZ, Pei HP, Watson DK, Papas TS (2000) EAP1/Daxx interacts with ETS1 and represses transcriptional activation of ETS1 target genes. Oncogene 19: 745-753.

49. Kim CA, Phillips ML, Kim W, Gingery M, Tran HH, et al. (2001) Polymerization of the SAM domain of TEL in leukemogenesis and transcriptional repression. EMBO J 20: 4173-4182.

50. Green SM, Coyne HJI, McIntosh LP, Graves BJ (2010) DNA binding by the ETS protein TEL (ETV6) is regulated by autoinhibition and self-association. J Biol Chem 285: 18496-18504.

51. Bojovic BB, Hassell JA (2001) The PEA3 Ets transcription factor comprises multiple domains that regulate transactivation and DNA binding. J Biol Chem 276: 4509-4521.

52. Greenall A, Willingham N, Cheung E, Boam DS, Sharrocks AD (2001) DNA binding by the ETS-domain transcription factor PEA3 is regulated by intramolecular and intermolecular protein.protein interactions. J Biol Chem 276: 16207-16215.

53. de Launoit Y, Baert JL, Chotteau A, Monte D, Defossez PA, et al. (1997) Structure-function relationships of the PEA3 group of Ets-related transcription factors. Biochem Mol Med 61: 127-135.

54. Sharrocks AD (2002) Complexities in ETS-domain transcription factor function and regulation: lessons from the TCF (ternary complex factor) subfamily. The Colworth Medal Lecture. Biochem Soc Trans 30: 1-9.

55. Graves BJ, Cowley DO, Goetz TL, Petersen JM, Jonsen MD, et al. (1998) Autoinhibition as a transcriptional regulatory mechanism. Cold Spring Harb Symp Quant Biol 63: 621-629.

56. Rabault B, Ghysdael J (1994) Calcium-induced phosphorylation of ETS1 inhibits its specific DNA binding activity. J Biol Chem 269: 28143-28151.

57. Cowley DO, Graves BJ (2000) Phosphorylation represses Ets-1 DNA binding by reinforcing autoinhibition. Genes Dev 14: 366-376.

58. Pufall MA, Lee GM, Nelson ML, Kang HS, Velyvis A, et al. (2005) Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. Science 309: 142-145.

59. Lee GM, Donaldson LW, Pufall MA, Kang HS, Pot I, et al. (2005) The structural and dynamic basis of Ets-1 DNA binding autoinhibition. J Biol Chem 280: 7088-7099.

60. Koizumi S, Fisher RJ, Fujiwara S, Jorcyk C, Bhat NK, et al. (1990) Isoforms of the human ets-1 protein: generation by alternative splicing and differential phosphorylation. Oncogene 5: 675-681.

61. Jorcyk CL, Watson DK, Mavrothalassitis GJ, Papas TS (1991) The human ETS1 gene: genomic structure, promoter characterization and alternative splicing. Oncogene 6: 523-532.

62. Grenningloh R, Miaw SC, Moisan J, Graves BJ, Ho IC (2008) Role of Ets-1 phosphorylation in the effector function of Th cells. Eur J Immunol 38: 1700-1705.

63. Goetz TL, Gu TL, Speck NA, Graves BJ (2000) Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. Mol Cell Biol 20: 81-90.

64. Behre G, Whitmarsh AJ, Coghlan MP, Hoang T, Carpenter CL, et al. (1999) c-Jun is a JNK-independent coactivator of the PU.1 transcription factor. J Biol Chem 274: 4939-4946.

65. Verger A, Buisine E, Carrere S, Wintjens R, Flourens A, et al. (2001) Identification of amino acid residues in the ETS transcription factor Erg that mediate Erg-Jun/Fos-DNA ternary complex formation. J Biol Chem 276: 17181-17189.

66. Bassuk AG, Leiden JM (1995) A direct physical association between ETS and AP-1 transcription factors in normal human T cells. Immunity 3: 223-237.

67. Batchelor AH, Piper DE, de la Brousse FC, McKnight SL, Wolberger C (1998) The structure of GABP alpha/beta: an ETS domain ankyrin repeat heterodimer bound to DNA. Science 279: 1037-1041.

68. Liu H, Holm M, Xie XQ, Wolf-Watz M, Grundstrom T (2004) AML1/Runx1 recruits calcineurin to regulate GM-CSF by Ets1 activation. J Biol Chem 279: 29398-29408.

69. Garvie CW, Hagman J, Wolberger C (2001) Structural studies of Ets-1/Pax5 complex formation on DNA. Mol Cell Biol 8: 1267-1276.

70. Lefstin JA, Yamamoto KR (1998) Allosteric effects of DNA on transcriptional regulators. Nature 392: 885-888.

71. Janknecht R, Nordheim A (1993) Gene regulation by Ets proteins. Biochim Biophys Acta 1155: 346-356.

72. Gille H, Kortenjann M, Thomae O, Moomaw C, Slaughter C, et al. (1995) ERK phosphorylation potentiates Elk-1 mediated ternary complex formation and transactivation. EMBO J 14: 951-962.

73. Li QJ, Yang SH, Maeda Y, Sladek FM, Sharrocks AD, et al. (2003) MAP kinase phosphorylation-dependent activation of Elk-1 leads to activation of the co-activator p300. EMBO J 22: 281-291.

74. Liu H, Grundstrom T (2002) Calcium regulation of GM-CSF by calmodulin-dependent kinase II phosphorylation of Ets1. Mol Biol Cell 13: 4497-4507.

75. Orlando V (2000) Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. Trends in Biochem Sci 25: 99-104.

76. Ren B, Robert F, Wyrick JJ, Aparicio O, Jennings EG, et al. (2000) Genome-wide location and function of DNA binding proteins. Science 290: 2306-2309.

77. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y, et al. (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nat Methods 4: 651-657.

78. Hollenhorst PC, Shah AA, Hopkins C, Graves BJ (2007) Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. Genes Dev 21: 1882-1894.

79. Boros J, O'Donnell A, Donaldson IJ, Kasza A, Zeef L, et al. (2009) Overlapping promoter targeting by Elk-1 and other divergent ETS-domain transcription factor family members. Nucleic Acids Res 37: 7368-7380.

80. Feng Y, Goulet AC, Nelson MA (2004) Identification and characterization of the human Cdc2l2 gene promoter. Gene 330: 75-84.

81. Carroll JS, Meyer CA, Song J, Li W, Geistlinger TR, et al. (2006) Genome-wide analysis of estrogen receptor binding sites. Nat Genet 38: 1289-1297.

82. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, et al. (2005) Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 122: 33-43.

83. Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, et al. (2006) A global map of p53 transcription-factor binding sites in the human genome. Cell 124: 207-219.

84. Rabinovich A, Jin VX, Rabinovich R, Xu X, Farnham PJ (2008) E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. Genome Res 18: 1763-1777.

85. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, et al. (2009) Diversity and complexity in DNA recognition by transcription factors. Science 324: 1720-1723.

86. Meijsing SH, Pufall MA, So AY, Bates DL, Chen L, et al. (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. Science 324: 407-410.

87. Lieb JD, Liu X, Botstein D, Brown PO (2001) Promoter-specific binding of Rap1 revealed by genome-wide maps of protein-DNA association. Nat Genet 28: 327-334.

88. Iyer VR, Horak CE, Scafe CS, Botstein D, Snyder M, et al. (2001) Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. Nature 409: 533-538.

89. Schodel J, Oikonomopoulos S, Ragoussis J, Pugh CW, Ratcliffe PJ, et al. (2011) High-resolution genome-wide mapping of HIF-binding sites by ChIP-seq. Blood 117: e207-217.

90. Rando OJ, Chang HY (2009) Genome-wide views of chromatin structure. Annu Rev Biochem 78: 245-271.

91. Valouev A, Johnson SM, Boyd SD, Smith CL, Fire AZ, et al. (2011) Determinants of nucleosome organization in primary human cells. Nature 474: 516-520.

92. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457: 854-858.

93. Verdin E (1991) DNase I-hypersensitive sites are associated with both long terminal repeats and with the intragenic enhancer of integrated human immunodeficiency virus type 1. J Virol 65: 6790-6799.

94. Shibata Y, Crawford GE (2009) Mapping regulatory elements by DNaseI hypersensitivity chip (DNase-Chip). Methods Mol Biol 556: 177-190.

95. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, et al. (2008) High-resolution mapping and characterization of open chromatin across the genome. Cell 132: 311-322.

96. Song L, Zhang Z, Grasfeder LL, Boyle AP, Giresi PG, et al. (2011) Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. Genome Res 21: 1757-1767.

97. Sterner DE, Berger SL (2000) Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev 64: 435-459.

98. Kouzarides T (2007) Chromatin modifications and their function. Cell 128: 693-705.

99. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, et al. (2005) A high-resolution map of active promoters in the human genome. Nature 436: 876-880.

100. Kim TH, Barrera LO, Qu C, Van Calcar S, Trinklein ND, et al. (2005) Direct isolation and identification of promoters in the human genome. Genome Res 15: 830-839.

101. Briggs SD, Bryk M, Strahl BD, Cheung WL, Davie JK, et al. (2001) Histone H3 lysine 4 methylation is mediated by Set1 and required for cell growth and rDNA silencing in Saccharomyces cerevisiae. Genes Dev 15: 3286-3295.

102. Ng HH, Robert F, Young RA, Struhl K (2003) Targeted recruitment of Set1 histone methylase by elongating Pol II provides a localized mark and memory of recent transcriptional activity. Mol Cell 11: 709-719.

103. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39: 311-318.

104. Roh TY, Cuddapah S, Cui K, Zhao K (2006) The genomic landscape of histone modifications in human T cells. Proc Natl Acad Sci U S A 103: 15782-15787.

105. Suzuki M, Yamada T, Kihara-Negishi F, Sakurai T, Hara E, et al. (2006) Site-specific DNA methylation by a complex of PU.1 and Dnmt3a/b. Oncogene 25: 2477-2488.

106. Lucas ME, Crider KS, Powell DR, Kapoor-Vazirani P, Vertino PM (2009) Methylation-sensitive regulation of TMS1/ASC by the Ets factor, GA-binding protein-alpha. J Biol Chem 284: 14698-14709.

107. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25: 1010-1022.

108. Erfurth FE, Popovic R, Grembecka J, Cierpicki T, Theisler C, et al. (2008) MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression. Proc Natl Acad Sci U S A 105: 7517-7522.

CHAPTER 2

DNA SPECIFICY DETERMINATES ASSOCIATE WITH DISTINCT TRANSCRIPTION FACTOR FUNCTIONS

OPEN ACCESS Freely available online PLoS GENETICS

DNA Specificity Determinants Associate with Distinct Transcription Factor Functions

Peter C. Hollenhorst [1], Katherine J. Chandler [1], Rachel L. Poulsen [2], W. Evan Johnson [1,2], Nancy A. Speck [3], Barbara J. Graves [1]*

1 Huntsman Cancer Institute, Department of Oncological Sciences, University of Utah, Salt Lake City, Utah, United States of America, 2 Department of Statistics, Brigham Young University, Provo, Utah, United States of America, 3 Abramson Family Cancer Research Institute, Department of Cell and Developmental Biology, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

Abstract

To elucidate how genomic sequences build transcriptional control networks, we need to understand the connection between DNA sequence and transcription factor binding and function. Binding predictions based solely on consensus predictions are limited, because a single factor can use degenerate sequence motifs and because related transcription factors often prefer identical sequences. The ETS family transcription factor, ETS1, exemplifies these challenges. Unexpected, redundant occupancy of ETS1 and other ETS proteins is observed at promoters of housekeeping genes in T cells due to common sequence preferences and the presence of strong consensus motifs. However, ETS1 exhibits a specific function in T cell activation; thus, unique transcriptional targets are predicted. To uncover the sequence motifs that mediate specific functions of ETS1, a genome-wide approach, chromatin immunoprecipitation coupled with high-throughput sequencing (ChIP-seq), identified both promoter and enhancer binding events in Jurkat T cells. A comparison with DNase I sensitivity both validated the dataset and also improved accuracy. Redundant occupancy of ETS1 with the ETS protein GABPA occurred primarily in promoters of housekeeping genes, whereas ETS1 specific occupancy occurred in the enhancers of T cell–specific genes. Two routes to ETS1 specificity were identified: an intrinsic preference of ETS1 for a variant of the ETS family consensus sequence and the presence of a composite sequence that can support cooperative binding with a RUNX transcription factor. Genome-wide occupancy of RUNX factors corroborated the importance of this partnership. Furthermore, genome-wide occupancy of co-activator CBP indicated tight co-localization with ETS1 at specific enhancers, but not redundant promoters. The distinct sequences associated with redundant versus specific ETS1 occupancy were predictive of promoter or enhancer location and the ontology of nearby genes. These findings demonstrate that diversity of DNA binding motifs may enable variable transcription factor function at different genomic sites.

Citation: Hollenhorst PC, Chandler KJ, Poulsen RL, Johnson WE, Speck NA, et al. (2009) DNA Specificity Determinants Associate with Distinct Transcription Factor Functions. PLoS Genet 5(12): e1000778. doi:10.1371/journal.pgen.1000778

Editor: Michael Snyder, Stanford University School of Medicine, United States of America

Received August 25, 2009; Accepted November 18, 2009; Published December 18, 2009

Funding: This work was supported by the National Institutes of Health (www.nih.gov) GM38663 (to BJG), HL089969 (to NAS and BJG), and CA42014 (to the Huntsman Cancer Institute for support of core facilities). BJG acknowledges funding from the Huntsman Cancer Institute (www.hci.utah.edu)/Huntsman Cancer Foundation (www.huntsmancancerfoundation.org). The funders had no role in study design, data collection and analysis decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: Barbara.Graves@hci.utah.edu

Introduction

Transcriptional regulation of gene expression is programmed through DNA sequence elements, termed promoters and enhancers. This genomic hard-wiring represents binding sites for transcription factors that have sequence specific DNA recognition and control development and homeostasis. Although the fundamental properties of protein-DNA recognition are well established, the advent of powerful technologies that provide genome-wide occupancy data has only recently allowed observation of these interactions in vivo. The emerging picture is that no single sequence motif fully explains all in vivo binding [1–3]. Furthermore, the in vitro derived consensus motifs are often present in only a minority of bound regions. These findings bring into question the purpose of binding site sequence variations. Possibilities are illustrated by experimental analysis of subsets of sites gathered from genomic data. For example, the PHA4/FOXO binding sites that program pharynx development in *C. elegans* differ in affinity, and thus carry developmental programming information dictating time of expression [4]. In yeast, PHO4 responsiveness to phosphate levels is regulated by alterative sequence motifs that affect affinity and program different roles for binding sites [5]. NF-κB and GR binding site variants can alter the repressing or activating transcriptional activity of the factor once it is bound [6,7]. The challenge of genomic databases is how to take full advantage of the vast number of binding sites, yet parse out functional consequences of variation. To realize their full potential, genomic approaches to transcriptional networks must go beyond a description of factor occupancy to include correlates of function.

We focus on the transcription factor ETS1 that provides a variety of contexts to address these central questions. ETS1 is a member of the ETS family of transcription factors that display similar DNA binding properties, including the recognition of a core GGA(A/T) motif. ETS family members are extensively co-expressed [8,9]. For example mRNA of 17 ets genes, including ETS1 , is present in Jurkat T cells. Despite overlapping expression

Author Summary

Genomes contain sequences that encode both gene products and the instructions for where and when each gene is expressed. This gene expression code is critical for normal development and goes awry in disease processes such as cancer. The gene expression code is interpreted by proteins called transcription factors that bind to particular DNA sequences and carry instructions for gene activation or repression. This recognition code is challenged by the presence of highly-similar transcription factors that prefer almost identical DNA sequences. In addition, studies in living cells indicate that individual transcription factors have significant flexibility in sequence recognition. Here, we identify thousands of positions in the genome of human T cells that are bound by the transcription factor ETS1. These data, along with comparisons to other genomic datasets, allow us to identify DNA sequences that specify ETS1 binding while excluding binding of other related transcription factors. Furthermore, we discover that ETS1 binds more than one sequence and that these sequence variants can predict distinct biological functions of ETS1. Thus, this work contributes to our understanding of the gene expression code by addressing both how a transcription factor can bind unique genomic locations and why a transcription factor binds multiple DNA sequences.

and sequence preferences, experimental data indicate that individual ETS proteins have unique biological functions [10–13]. For ETS1, mouse deletion studies indicate a critical role in T cell activation [14]. This specific genetic function implies that ETS1 has a unique mechanism that allows it, but not other ETS proteins, to bind the promoters or enhancers of genes important for T cell activation. Finally, ETS1 functions, in part, by recruitment of the co-activator CBP to transcriptional control regions, presumably functioning to activate genes at which it binds [15–17]. We utilize this in depth understanding of ETS1 at both a biochemical and biological level to inform our genomic approach and facilitate functional analysis.

Initial genomic occupancy studies with ETS1 have provided insight into the genomic dilemma and led to unexpected observations. We previously identified two modes of ETS protein targeting to promoters in Jurkat T cells using chromatin immunoprecipitation and promoter microarrays (ChIP-chip). The surprising mode is redundant occupancy in which a sequence with the consensus CCGGAAGT is associated with occupancy of three different ETS transcription factors: ETS1, GABPA (GABP a), and ELF1. Because this sequence is consistent with the in vitro-derived consensus sequences derived from multiple ETS family members, we concluded that it can alternately recruit various ETS transcription factors. This redundant mode of binding generally occurs in the promoters of housekeeping genes and may represent shared function of the ETS family in the maintenance of constitutive expression. The second ETS binding mode is specific occupancy (e.g. ETS1, but not GABPA or ELF1), which requires a GGA core motif, but is not associated with a close match to the consensus ETS sequence. We proposed that specific targets would mediate the specific biological functions of each ETS transcription factor. However, the promoter-limited approach did not identify a significant correlation between the specific targets of ETS1 and genes important for the role of ETS1 in T cell activation. Full investigation of this provocative dual role of the ETS family required an expansion to full genome analysis.

In this study we identified regions across the entire human genome occupied by ETS1, a DNA binding partner RUNX, and co-activator CBP in Jurkat T cells to decipher sequence determinants and investigate the biological significance of sequence diversity. We discovered a previously undescribed role for ETS1 at a large number of enhancers. Enhancer occupancy of ETS1 was associated with a unique variant of the ETS binding site and in vitro DNA binding assays illustrated how this variant sequence functions as an ETS1 specificity determinant. Enhancers co-occupied by ETS1 and RUNX contained a variant ETS sequence closely juxtaposed to a RUNX binding site – a composite sequence identical to that found in the T cell receptor enhancers. These distinct enhancer sequences contrasted with prior observations of sequences at ETS1 bound promoters. Importantly, ETS1 bound regions that contained the ETS/RUNX composite sequence were near genes important for T cell activation, thus establishing a tissue-specific, genomic dataset for a factor partnership. Furthermore, ETS1 was closely associated with CBP occupancy at ETS1 specific enhancers, but not at redundantly occupied promoters. By using genomic datasets for DNA binding factors, in addition to correlates of DNase I sensitive regions, histone marks, and co-factor binding, we decoded the functionality of in vivo binding sequences.

Results

ETS1 and GABPA co-occupy active promoters, but ETS1 specifically occupies T cell enhancers

High-throughput sequencing coupled with chromatin immunoprecipitation (ChIP-seq) facilitates genome-wide searches for transcription factor binding sites [1]. We detected 19,420 bound regions at an empirical false discovery rate of 0.01 for ETS1 in Jurkat T cells using this approach. This included almost all (94%) of the 1086 ETS1 bound promoters previously identified by ChIP-chip [18] plus an additional 6116 promoters, indicating a potential for higher sensitivity. ETS1 bound promoter regions centered within 500 bp of a transcription start site (TSS) (Figure S1); therefore, a 500-bp limit was used for promoter definition. A large number of regions, 12,283, were not in promoters. We sought to establish the validity of these potential enhancer regions by comparison to other types of genome-wide datasets. One powerful dataset from primary $CD4^+$ T cells (thus comparable to the CD4+ Jurkat cell line) identifies DNase I accessible regions as mapped by high-throughput sequencing and ChIP-chip [19]. Based on the long history of linkage of DNase I sensitivity to enhancers, we screened ETS1 bound regions for overlap. 76% of ETS1 occupied regions overlapped with DNase I sensitivity. (Overlap was 98% for sites proximal to a TSS and 64% for distal sites.) This represents a significant enrichment over the mean 4% overlap with datasets randomly derived from control sequences ($P < 0.001$, Figure 1A). Randomly selected DNase I sensitive, ETS1 bound regions were verified by quantitative PCR as ETS1 occupied (13 of 15 ETS1 bound, Figure S2), whereas regions that were not DNase I sensitive included many apparent false positives (0 of 8 ETS1 bound, Figure S2). This strong correlation not only helped validate the ETS1 data, but also suggested that DNase I sensitivity is a strong correlate of robust ChIP signals. We proposed that the 14,824 ETS1 bound regions that overlap DNase I sensitive regions represent functional regions, and only these were considered in further analysis.

Datasets for histone marks in primary CD4+ T cells [20] also provide a measure of the activity of promoters and enhancers and test for relevance of factor binding. For example, H3K4 trimethylation correlates with active promoters [21]. 86% of ETS1

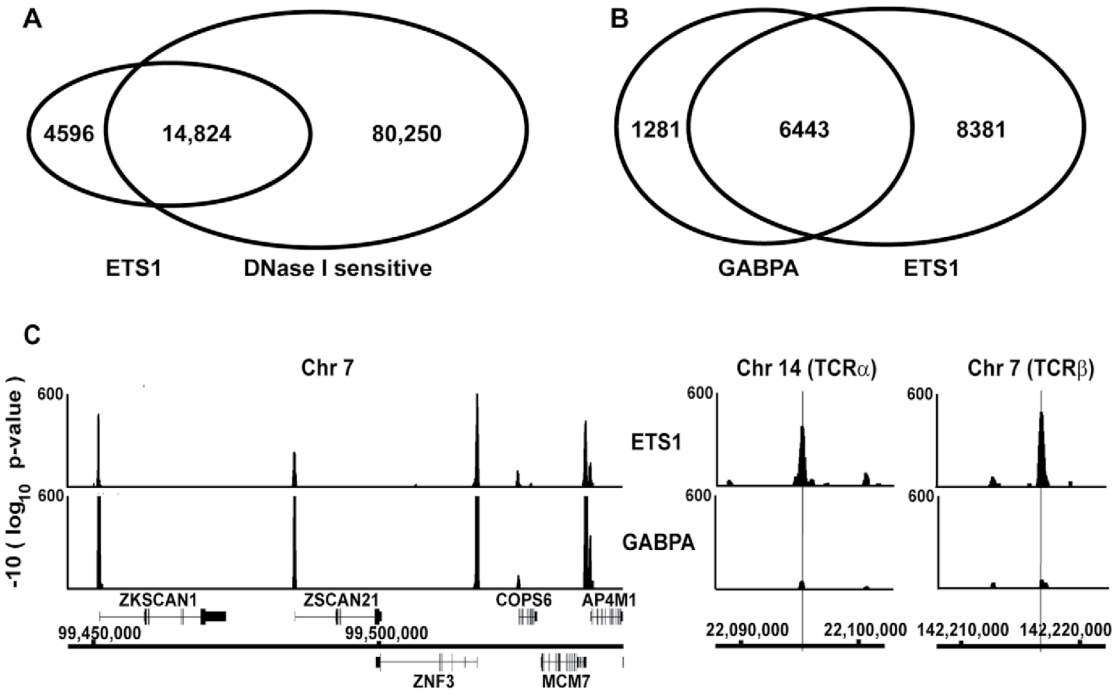


Figure 1. Genomic occupancy of ETS1 overlaps with DNase I sensitivity and GABPA occupancy. (A) Overlap of genomic regions bound by ETS1 in Jurkat T cells and regions found to be DNase I sensitive in CD4$^+$ T cells [19]. (B) Overlap of regions occupied by ETS1 and GABPA in Jurkat T cells [23]. Only regions that overlap with DNase I sensitivity were included. (C) Log transformed P values of ETS1 and GABPA occupancy in a scanning 250 bp window mapped using the Integrated Genome Browser (http://igb.bioviz.org/) to regions of the human chromosome (Chr) indicated by chromosomal coordinates (NCBI Build 36.1). Positions of Refseq genes are shown with genes transcribed from left to right above the nucleotide position bar and genes in the opposite orientation below. Vertical lines (right) indicate the ETS/RUNX binding sequences previously tested for function [27,44] in the TCRa enhancer (sequence GAGGATGTGGC) or the TCRb enhancer (sequence CAGGATGTGGT).
doi:10.1371/journal.pgen.1000778.g001

bound promoters had H3K4 tri-methylation compared to 28% of promoters without ETS1 (P, 0.0001) (Table 1). Likewise, H3K4 mono-methylation is enriched in enhancers [21,22]. For DNase I sensitive regions distal to a TSS, 52% with ETS1, but only 18% without ETS1, had an H3K4 mono-methyl mark (P, 0.0001) (Table 1). Therefore, ETS1 occupancy is enriched at regions with histone marks that are indicative of enhancers and active promoters.

Contrary to the expectation that family members with unique genetic functions would have exclusive binding sites, we previously discovered the majority of proximal promoters bound by ETS1, are also occupied redundantly by other ETS proteins (e.g. GABPA and ELF1). The small number of ETS1 specific sites identified limited the robustness and fruitfulness off urther analysis of this class of targets [18]. Using a much larger ChIP-seq dataset that

Table 1. Marks of enhancers and active promoters are associated with ETS1 occupancy.

Category [a]	Number [b]	H3K4me3 [cd]	H3K4me1 [cd]	CBP [d]
Promoters occupied by ETS1	8236	86%	16%	75%
Promoters lacking ETS1	14,445	28%	11%	9%
DNase I sensitive promoters lacking ETS1	5961	58%	19%	20%
Distal DNase I sensitive regions occupied by ETS1	7501	46%	52%	68%
Distal DNase I sensitive regions lacking ETS1	77,141	10%	18%	5%

a Promoters are defined as 500 bp upstream and downstream of a RefSeq TSS. Distal regions are regions with the center greater than 500 bp from a RefSeq TSS. ETS1 occupied regions contain the center of an ETS1 bound region within the promoter or distal region. Totals differ from the number of ETS1 bound regions (Figure 1) because in some cases multiple promoters or DNase I sensitive regions overlap with a single ETS1 bound region.
b The number of regions in the category.
c Positions of histone marks in CD4$^+$ T cells were determined by analyzing published ChIP-seq reads [20] with the Useq bioinformatics package.
d The percentage of regions in the category containing the indicated histone mark or CBP occupancy.
doi:10.1371/journal.pgen.1000778.t001

includes enhancer regions, we were poised to identify targets that mediate specific functions of ETS1. However, we first had to identify which of the 14,824 ETS1 bound regions were redundant sites, thus not strong candidates to mediate specific functions. We analyzed genome-wide occupancy data reported for GABPA in Jurkat cells [23] with the same methodology as our ETS1 analysis. There were 7724 GABPA bound regions, of which 6443 were redundantly bound by ETS1 (Figure 1B), illustrated graphically on a genome section in Figure 1C (left). The remaining 8381 ETS1 specific regions were exemplified by the T cell receptor (TCR) a and b enhancer loci (Figure 1C, right), which display ETS1, not GABPA, occupancy. 67% of GABPA and ETS1 co-occupied regions were proximal to a TSS consistent with previous findings that discovered this dominant class of promoter binding events [18]. In contrast, 68% of ETS1 specific regions were distal to the nearest TSS indicating enhancer regions and are candidates to mediate the specific functions of ETS1.

Genetic experiments have implicated ETS1 in T cell activation; therefore, we predicted that genomic sites that specifically bind ETS1 should be associated with genes necessary for T cell function. The genes nearest to distal ETS1 bound regions were assumed to be a reasonable estimate of ETS1 regulated genes. This gene set was compared to genes with 20-fold higher mRNA expression levels in CD4$^+$ T cells than the median expression in multiple cell types (287 genes) and, as a control, to 268 genes specific to pancreas (Figure 2). Compared to all genes, or pancreas specific genes, genes that displayed T cell–specific expression were more likely to be near one or more distal ETS1 bound regions. Furthermore, as the number of nearby distal ETS1 bound regions increased, the difference between the T cell–specific categories and the control categories became more apparent. Thus, distal ETS1 binding was associated with a tissue-specific role of ETS1 in T cells, further validating the functionality of ETS1 specific regions.

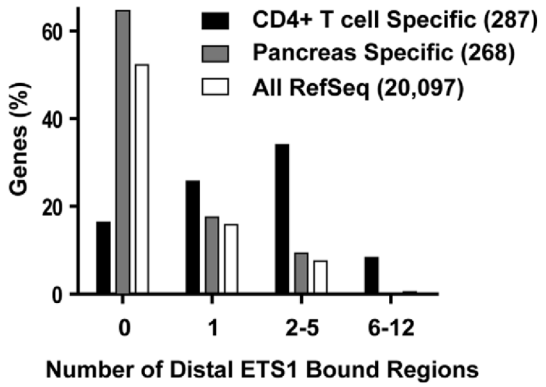


Figure 2. Distal ETS1 bound regions are found near T cell–specific genes. The frequency of neighboring, distal, ETS1 bound regions was compared for three categories of genes; all Refseq genes, CD4$^+$ T cell–specific genes, and pancreas specific genes. Tissue specific gene lists were derived from the GNF SymAtlas database [52] and were based on the level of mRNA in T cells or pancreas compared to the median in all surveyed tissues with cutoffs (20-fold higher for T cells, 5-fold higher for pancreas) that returned similar sized lists. The number of genes in each category is indicated in parenthesis. Each ETS1 bound region was matched to a single gene based on the nearest RefSeq TSS. The percent of genes in each category associated with one or more distal ETS1 bound regions (greater than 500 bp from the TSS) was plotted.
doi:10.1371/journal.pgen.1000778.g002

ETS1 occupancy of enhancers was associated with a distinct sequence

At this point we had a dataset of ETS1 bound regions that tracked with distal enhancer marks and T cell function. We sought to determine whether such bound regions displayed a unique sequence that would be responsible for ETS1 specific binding. Unbiased searching for overrepresented sequences was performed with the MEME algorithm [24]. ETS1 bound regions were grouped according to proximity to the nearest TSS (proximal versus distal) and specificity (redundant: overlap with GABPA; specific: no overlap with GABPA) to provide experimental and control datasets. The most over-represented sequence motifs in redundant, proximal regions (Motif 1) and specific, proximal regions (Motif 2) were identical to the motifs previously identified in redundant and specific promoter proximal regions [18]. In contrast, analysis of distal, specific ETS1 bound regions identified a third, distinct motif (Motif 3) as the most over-represented (Figure 3). The major differences from the ETS family consensus (CCGGAAGT) present in Motif 1 were the almost exclusive presence of an A at the second position and the inclusion, in some instances, of a T at the sixth position (CAGGA(A/T)GT). Therefore, specific ETS1 binding to enhancers is associated with a sequence distinct from those found at ETS1 bound promoters.

A more directed bioinformatics approach assessed the importance of single nucleotide changes from the ETS consensus for enhancer and promoter occupancy of ETS1. ETS1 bound regions were partitioned into equally-sized sets of ETS1/GABPA redundantly occupied promoters and ETS1 specific enhancers. The number of occurrences of 8-mer sequences was reported relative to the number of occurrences expected in a set of equally-sized random sequences (Table 2). The ETS consensus sequence (CCGGAAGT) was enriched in redundant promoters, but not in ETS1 specific enhancers. The enrichment of every possible single nucleotide change to the ETS consensus was then determined. Only the change of the C at the second position to an A (CaGGAAGT) resulted in a significant enrichment ($P < 0.0001$) in ETS1 specific enhancers. However, this sequence was also enriched in redundant promoters. The A to T change at the sixth position (CCGGAtGT) reduced significantly the enrichment in redundant proximal regions ($P < 0.0001$), but not distal regions ($P = 0.3$). Furthermore, the combination of both nucleotide changes (CaGGAtGT) resulted in enrichment at ETS1 specific enhancers ($P < 0.0001$), but not redundant promoters ($P = 0.3$). Strikingly, a change of only two nucleotides inverted the enrichment pattern at redundant promoters and ETS1 specific enhancers. We concluded that the sequence CAGGATGT is a specificity element for ETS1.

In considering specificity within families of transcription factors the preference for a particular sequence may be due to intrinsic DNA binding properties of different family members. To test the ability of the two nucleotide changes to act alone or in combination to select for ETS1 we measured the relative binding affinity of ETS1 and second ETS factor, ELF1 which is also reported to be active in T cells [25]. Indeed, ELF1 is present at redundant, but not ETS1 specific promoters in Jurkat T cells in a similar manner to GABPA [18]. Binding affinity for an ETS consensus sequence, each single nucleotide variant, and the two nucleotide variant was interrogated *in vitro* with purified proteins by quantitative gel shift (Figure 4). The A to T change resulted in a loss of affinity for ELF1 (3.6-fold loss), but not for ETS1. The C to A change showed a modest effect on affinity and no discrimination between ELF1 and ETS1 (1.9-fold versus 1.5-fold loss). In contrast, the change of both nucleotides caused an 18.3-fold loss of affinity for ELF1, but only a 2.4-fold loss for ETS1. We

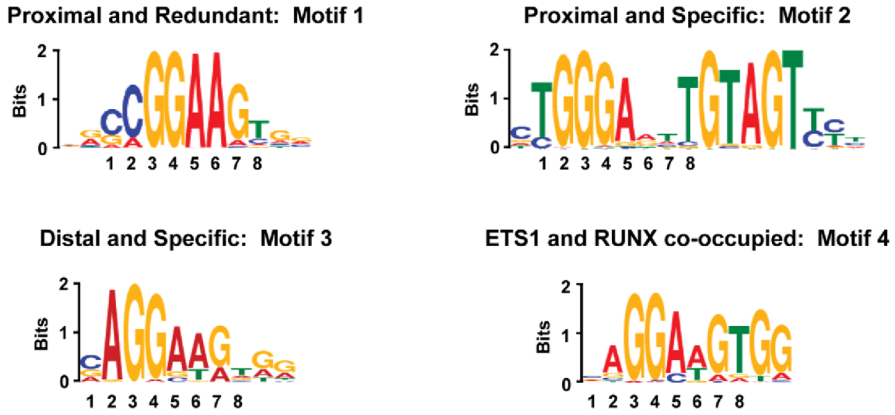


Figure 3. Distinct sequence motifs are over-represented in different subsets of ETS1 bound regions. The indicated subsets of ETS1 bound regions were rank ordered by log transformed binomial P value and the top 250 regions were searched for over-represented sequences by MEME [24]. The most over-represented position weight matrix for each subset is represented (E-values: Motif 1, $9.9 \times 10^{2\ 281}$; Motif 2, $2.8 \times 10^{2\ 287}$; Motif 3, $8.8 \times 10^{2\ 105}$; Motif 4, $1.8 \times 10^{2\ 142}$). The height of each nucleotide indicates conservation at that position. Eight nucleotide positions in ETS binding sites are numbered for reference. ETS bound regions were classified either proximal (center of region within 500 bp of a TSS) or distal (center of region greater than 500 bp from a TSS). ETS1 bound regions were classified as redundant if they overlapped with a GABPA bound region and specific if they did not. ETS1 bound regions were classified as RUNX co-occupied if they overlapped with a RUNX bound region.
doi:10.1371/journal.pgen.1000778.g003

concluded that the two nucleotide variant sequence CaGGAtGT serves as a specificity determinant for ETS1 versus ELF1 due to an intrinsic binding property of ETS1.

RUNX co-occupies enhancers with ETS1 through a distinct sequence motif

Unique cooperative DNA binding between closely apposed binding proteins can also drive specific occupancy of transcription factors. A well-characterized partnership for ETS1 is with the RUNX factors [26,27]. Interestingly, the consensus derived from the most frequent nucleotides at each position of Motif 3 (CAGGAAGTGG) is similar to the sequences at the TCR b (CAGGATGTGG) and TCR a (GAGGATGTGG) enhancers that support cooperative binding of ETS1 with RUNX1 through an ETS/RUNX composite sequence (RUNX consensus YGYGGY). To test whether ETS1 bound enhancers were co-occupied by RUNX factors, genome-wide occupancy of RUNX1/3 (RUNX) was determined. Again, only regions that co-localized with DNase I sensitivity were considered bound. Strikingly, 64% of the 1075 RUNX bound regions were co-occupied by ETS1. In contrast, only 14% of RUNX bound regions were co-occupied by GABPA (Figure 5A). 77% of the ETS1/RUNX co-occupied regions were ETS1 specific and distal to a TSS (compared to 37% of ETS1 bound regions lacking RUNX), suggesting a role in T cell enhancer function. An unbiased search with MEME for overrepresented sequence motifs in regions co-occupied by ETS1 and RUNX identified a motif (Motif 4) similar to Motif 3, but with the RUNX binding site more strongly represented (Figure 3). Like many sequence identification algorithms, MEME is biased towards strongly preferred spacing distances between two binding sites. To test whether other spacings of ETS and RUNX sites were also over-represented in ETS/RUNX bound regions, the distance from each ETS sequence to the nearest RUNX sequence was plotted (Figure 5B). This analysis indicated that only the spacing found by MEME was over-represented in these regions. Therefore, ETS1 and RUNX co-occupy enhancer regions in T cells through a composite ETS/RUNX binding site similar to those found in the T cell receptor enhancers. These findings indicate that pairing with a neighboring DNA binding motif, in conjunction with intrinsic DNA binding properties, can drive specificity.

Distinct sequence motifs are correlated with unique modes of ETS1 occupancy

ETS1 occupancy of enhancers is associated with T cell–specific genes (Figure 2) and ETS1 specific motifs (Motifs 3 and 4), whereas promoter occupancy is associated with housekeeping genes [18] and shows enrichment for sequences (Motif 1) that cannot distinguish family members. The value of these motifs will be in their predictive accuracy. All ETS1 bound regions were searched for Motif 1 and Motif 4 with PATSER [28]. Regions containing Motif 1 were more likely to be found in promoters, and regions containing Motif 4 were more likely to be found in enhancers (Table 3, Table S1). Associated genes, as determined by nearest TSS, were searched for over-represented ontologies with the GoMiner program [29]. Genes with Motif 4 were associated with T cell activation categories, whereas those with Motif 1 were associated with housekeeping ontologies (Table 3). Therefore, each motif is predictive of the type of transcriptional control element and class of ETS1 target gene.

CBP/p300 co-localizes with ETS1 at enhancers, but not at promoters

The emerging differences for ETS1 at promoters versus enhancers opened the possibility of distinct functions of ETS1 at these loci. One mechanism of transcriptional activation by ETS1 is the recruitment of the co-activators CBP and p300 [16]. Identification of p300 occupancy within the 30 mb ENCODE region of the human genome revealed a greater proportion at distal sites than at promoters [21], and p300 has been shown to mark tissue specific enhancers in mice [30]. Thus, we proposed that ETS1 would recruit CBP/p300 to T cell–specific enhancers, but not promoters. Genome-wide occupancy for CBP detected 14,374 CBP bound/DNase I sensitive regions in Jurkat T cells.

Table 2. ETS1 occupancy of redundant promoters and specific enhancers is associated with distinct sequences.

Sequence [a]	Enrichment in redundant and proximal regions [b]	Enrichment in specific and distal regions [b]
CCGGAAGT	33	1
gCGGAAGT	15	1
aCGGAAGT	4	1
tCGGAAGT	2	0
CgGGAAGT	2	1
CaGGAAGT	10	14
CtGGAAGT	2	2
CCcGAAGT	1	0
CCaGAAGT	1	2
CCtGAAGT	1	1
CCGcAAGT	0	1
CCGaAAGT	0	0
CCGtAAGT	1	0
CCGGgAGT	2	0
CCGGcAGT	2	0
CCGGtAGT	1	0
CCGGAgGT	2	1
CCGGAcGT	2	0
CCGGAtGT	4	1
CCGGAAcT	4	0
CCGGAAaT	5	1
CCGGAAtT	1	0
CCGGAAGg	9	1
CCGGAAGc	23	1
CCGGAAGa	8	1
CaGGAtGT	1	10

a The ETS consensus from Motif 1 (Figure 3), every possible single nucleotide change, and one double nucleotide change are shown.
b The number of occurrences of each octamer in equally-sized data sets of ETS1 bound, redundant, proximal regions and ETS1 bound, specific, distal regions were counted. Enrichment was determined by dividing this number by the expected number of occurrences of an octamer in random sequence space of the same size and rounding to the nearest integer. All enrichment values greater than two are significant, P, 0.0001, Fisher's exact test.
doi:10.1371/journal.pgen.1000778.t002

CBP bound regions showed a surprisingly high overlap with ETS1 bound regions at both redundant promoters (75%, P, 0.0001) and ETS1 occupied enhancers (68%, P, 0.0001) compared to regions not bound by ETS1 (Table 1). The strong presence of CBP corroborated the general functionality of ETS1 binding sites.

Due to the unexpected equivalence of CBP overlap at both enhancers and promoters we investigated the connection between CBP binding and ETS1 function by a more detailed mapping method that presented the two types of sites as a class average (Figure 6A). At redundantly occupied promoters, the location of ETS1, CBP, GABPA, H3K4 tri-methylation, and Motif 1 were plotted relative to the TSS. At ETS1 occupied enhancers the location of CBP, RUNX, H3K4 mono-methylation, and Motif 4 were plotted relative to the center of the ETS1 bound region. At promoters ETS1 and GABPA binding were coincident with the consensus ETS binding site at a position 25–30 bp upstream of the transcription start site. This extremely TSS proximal location and the location of histone H3K4 tri-methylation on either side of the ETS1 bound region indicated that redundant ETS binding occurs in the nucleosome-free region [31]. CBP occupancy was co-incident with the downstream H3K4 tri-methyl, but not ETS1 and GABPA binding, suggesting that CBP is not directly bound by ETS factors at promoters. In contrast, at enhancers, CBP, ETS1, and RUNX binding overlapped, suggesting that ETS1 and/or RUNX may directly bind CBP. Again, ETS1 occupied a region between histone marks, in this case H3K4 mono-methyl, indicating that ETS1 binds between nucleosomes.

To test directly whether ETS1 was necessary for occupancy of RUNX and CBP, ETS1 protein levels were knocked-down by two independent shRNAs and occupancy was monitored by ChIP (Figure 6B and 6C). Decreased ETS1 protein levels correlated with a loss of ETS1, CBP, and RUNX ChIP enrichment at the TCR b enhancer (containing Motif 4). We concluded that ETS1 is critical for recruitment or stable binding of CBP in enhancers important for T cell activation. In contrast, reduction of ETS1 occupancy in a redundantly occupied promoter (containing Motif 1) did not affect CBP enrichment. Thus, distinct sequence motifs at ETS1 binding sites correlate with not only different types of regulatory elements, but also distinct histone marks and co-activator binding. We conclude that these sequences mediate unique functions of ETS1.

Discussion

Comprehensive identification of ETS1 binding sites in Jurkat T cells revealed that ETS1 was present at a large number of enhancers in a context distinct from that previously observed at promoters. Enhancers differed from promoters in the sequence elements that bind ETS1, in the compliment of neighboring proteins and histone marks, and in the ontology of nearby genes. Specifically, the tissue specific function of ETS1 correlated with enhancer binding via an ETS/RUNX composite sequence. Furthermore, our data indicated that this T cell–specific enhancer function, but not the housekeeping promoter function of ETS1, is associated with co-localization of the co-activator CBP. Therefore, the sequences motifs identified here are associated with specific enhancer occupancy of ETS1 and define a different function than the sequence associated with redundant ETS1 occupancy at promoters.

Overlap with DNase I sensitivity improves the accuracy of a ChIP-seq dataset

Selecting regions at which ChIP-seq enrichment coincided with DNase I sensitivity improved the accuracy of a dataset of transcription factor bound regions. The fraction of regions removed (24% of ETS1 bound regions, 39% of CBP bound regions and 70% of RUNX regions) may reflect the quality of the antibodies used for ChIP-seq. Removed regions had ETS1 binding properties (presence of binding motifs, correlations with other factors and histone marks), but at markedly lower levels than retained regions. Thus, we propose that the use of DNase I sensitivity screening improves the quality of a ChIP-seq dataset and may be particularly useful for the interpretation of data generated with suboptimal antibody reagents.

ETS1 binding determinants vary at specific enhancers and redundant promoters

The genome-wide set of ETS1 binding sites showed sequence variants that distinguish enhancer versus promoter binding events. Specific ETS1 occupancy of enhancers was associated with a sequence that varies by two nucleotides from the ETS consensus sequence used for redundant binding at promoters. Our bioinfor-

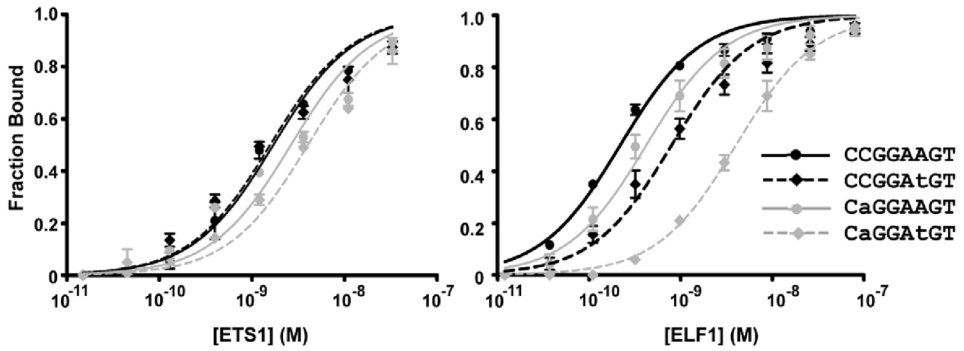


Figure 4. Intrinsic DNA binding affinity differences for ETS1 and ELF1. Full-length recombinant versions of the human ETS proteins ETS1 and ELF1 were purified from bacteria and assayed for affinity to radiolabeled oligonucleotides by gel-shift analysis. Datapoints represent the mean and standard error of the mean of two replicates for ETS1 and three replicates for ELF1. Each K_D was derived by curve fitting by nonlinear least-squares analysis of equilibrium binding curves with fraction of DNA bound = $1/(1+ K_D/[ETS1])$. The K_D of ETS1 and ELF1 for the ETS consensus sequence was $1.76\ 10^{-9}$ and $1.76\ 10^{-10}$ M, respectively. The fold decrease in affinity due to the A to T change, the C to A change, and the combination of both changes were 1.0, 1.5, and 2.4 for ETS1 and 3.6, 1.9, and 18.3 for ELF1, respectively.
doi:10.1371/journal.pgen.1000778.g004

matics analysis indicated that these two nucleotide changes are not equivalent (Table 2). The C to A change at the second position appeared to be required for specific ETS1 binding to enhancers, but also occurs at redundant promoters. In contrast, the A to T change at the sixth position appeared to restrict occupancy of redundant promoters, but not specific enhancers. The A to T change has previously been shown to provide specificity for ETS1 versus the ETS protein ELF1 in vitro [32]. Our in vitro comparison confirmed the role of this single nucleotide change and identified a dramatic specificity difference between ETS1 and ELF1 when both nucleotides were changed (Figure 4). However, the in vitro data did not explain why the C to A change alone appeared necessary for genomic enrichment in ETS1 specific enhancers (Table 2). Therefore, the nucleotide preferences at these ETS binding sequences are likely due to a combination of the intrinsic differences in DNA binding attributes of ETS proteins and other in vivo factors.

A striking difference between the ETS consensus sequence, CCGGAAGT, and the C to A variant, CaGGAAGT, is the susceptibility to DNA methylation. Indeed, methylation of this CpG dinucleotide within the consensus has been shown to block the binding of ETS proteins [33,34]. We have previously observed a very strong correlation between redundant ETS occupancy of promoters, the sequence CCGGAAGT and CpG islands [18]. The CpG islands at housekeeping promoters are generally hypomethylated, whereas CpG dinucleotides distributed in lower density throughout the genome are likely to be methylated [35,36]. Thus, ETS sites may be shielded from methylation at CpG island-

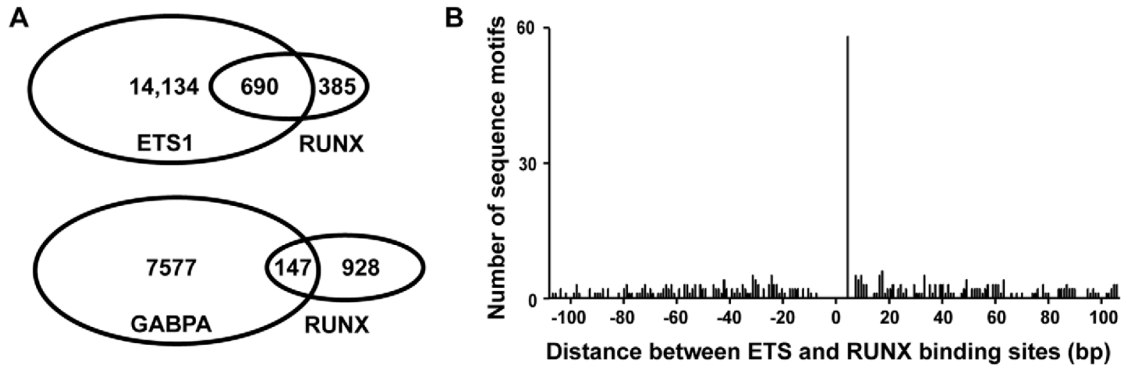


Figure 5. ETS1/RUNX co-occupancy correlates with specific spacing of ETS and RUNX binding sites. (A) Overlap of ETS1/RUNX and GABPA/RUNX bound regions. The RUNX antibody was raised against the conserved DNA binding domain and does not differentiate between the homologous RUNX1 and RUNX3 proteins present in T cells (N. Speck unpublished observation). (B) Spacing of ETS and RUNX binding sites in ETS1/RUNX co-occupied regions. The 690 ETS1 bound regions that were co-occupied by RUNX were scanned for matches to the in vitro derived position weight matrixes for ETS1 (M00032) and RUNX (M00271) from the Transfac database (http://www.biobase-international.com/index.php?id = transfac). For each ETS1 sequence found, the distance to all forward oriented RUNX sequences in the same region were determined such that a RUNX site 59 to an ETS site in the orientation CCGGAAGT was negative and 39 was positive. A similar mapping of RUNX sites in the reverse orientation returned no spacing frequencies higher than five. The prominent peak at a spacing of −5 bp correlates with the spacing and orientation in the composite sequence CAGGATGTGGT, from Motif 4 (Figure 3).
doi:10.1371/journal.pgen.1000778.g005

Table 3. ETS1 bound regions containing Motif 1 and Motif 4 have different characteristics.

Motif	Regions [a]	Promoter [b]	Most overrepresented ontologies [c]	*P* value [d]
4	1364	24%	Leukocyte activation	$7.66\ 10^{-10}$
			Phosphate metabolic process	$9.96\ 10^{-9}$
			Cell activation	$1.46\ 10^{-8}$
			Lymphocyte activation	$3.16\ 10^{-8}$
			Protein amino acid phosphorylation	$3.76\ 10^{-8}$
			T cell activation	$1.26\ 10^{-7}$
			Positive regulation of lymphocyte activation	$1.36\ 10^{-7}$
			Immune system process	$1.96\ 10^{-7}$
1	4492	59%	Macromolecule metabolic process	$4.96\ 10^{-29}$
			Biopolymer metabolic process	$1.56\ 10^{-26}$
			Primary metabolic process	$1.16\ 10^{-20}$
			Cellular metabolic process	$1.76\ 10^{-20}$
			Gene Expression	$2.36\ 10^{-19}$
			Nucleotide and nucleic acid metabolism	$2.86\ 10^{-18}$
			Metabolic process	$3.76\ 10^{-17}$
			RNA processing	$7.36\ 10^{-16}$

a Number of ETS1 bound regions containing a match to the PWM for Motif 1 and 4 as shown in Figure 3. PWM nucleotide frequencies and cutoffs are detailed in the Materials and Methods and Table S2.
b Percentage of regions with the center within 500 bp of a RefSeq TSS.
c Regions were mapped to the nearest RefSeq gene and gene lists were analyzed by GoMiner. Overrepresented ontologies are listed in the order returned by GoMiner with no editing.
d P value for each ontology category from GoMiner.
doi:10.1371/journal.pgen.1000778.t003

containing promoters, but not at enhancers. Therefore, the C to A change in the ETS binding sequences at enhancers may have evolved to protect these sites from the repressive effects of DNA methylation. Because other transcription factors whose binding sites bear a CpG dinucleotide (NRF-1, BoxA, SP1, CRE, and E-box) are also enriched in housekeeping promoters [37], the use of an alternate sequence in tissue-specific enhancers may also extend to these transcription factor families.

Another factor that might influence the ETS1 binding sequence observed *in vivo* is the presence of closely juxtaposed binding sites for other transcription factors. A subset of the ETS1 specific enhancers were co-occupied by RUNX and had a composite ETS/RUNX binding sequence (Motif 4). In the context of this sequence, the A to T change at the sixth position of the ETS sequence allows the RUNX sequence to be a closer match to the RUNX consensus (YGYGGT, sixth position underlined). This factor could contribute to the enrichment of the A to T change in specific enhancers. However, a T at the sixth position was no more likely in regions co-occupied by ETS1 and RUNX (Motif 4) than at ETS1 specific enhancers in general (Motif 3). This indicates that either an A or a T at this position can support ETS1 and RUNX co-occupancy. Furthermore, only 55% of the ETS1 bound regions that had the sequence CAGGATGT, had the full ETS/RUNX composite CAG-GATGTGG. We propose that the remaining 45% of regions recruit ETS1 either only through the ETS binding site, or in cooperation with other unidentified transcription factors. In conclusion, the sequences associated with ETS1 specific occupancy of enhancers reflect intrinsic differences in DNA binding or interactions with other factors and may be influenced by susceptibility to DNA methylation.

ETS1 and RUNX co-occupancy predict tissue-specific enhancers

Mice with an ETS1 gene disruption have reduced numbers of NK and NKT cells and show defects in T cell activation [14,38,39]. RUNX genes are essential for NK, NKT, and T cell differentiation [40–42]. However, the role of ETS1 and RUNX in these immune functions has not been fully understood on the level of individual target genes. Our data suggest that ETS1 and RUNX regulate genes important for T cell activation pathways by direct occupancy of nearby enhancers via a particular ETS/RUNX binding site. The gene categories presented in Table 3 suggest that the primary role of these transcription factors is not the direct activation of genes downstream of T cell receptor signaling, but rather the control of expression of the signaling machinery. This conclusion is consistent with the finding that ETS1 null T cells are defective in activation upon receptor stimulation, but respond normally to pharmacological stimulation, which bypasses membrane proximal signaling events [39].

The strongest determinant of ETS1 specificity, Motif 4, fixed the sequence of both ETS1 and RUNX binding sites as well as spacing and relative orientation of the two sites (Figure 3 and Figure 5). This strict conservation was somewhat surprising because ETS1 and RUNX1 bind to DNA cooperatively *in vitro* at a variety of other spacings and orientations [26,43]. Alternate spacing can also function in transcription activation *in vivo*. For example, the MMLV enhancer is activated by ETS1 and RUNX1 at a sequence in which the ETS and RUNX sites are four nucleotides further apart than in Motif 4 [44]. Furthermore, Motif 2 can also support ETS1 and RUNX1 cooperativity *in vitro* [18]. Motif 2 utilizes much more divergent ETS and RUNX sequences set two nucleotides further apart than in Motif 4. (Our analysis in Figure 5 does not identify this spacing because these sequences are too divergent from the Transfac ETS1 and RUNX motifs.) However, only Motif 4, not Motif 2 or the MMLV enhancer motif, was associated with ontologies aligned with T cell–specific functions (Table 3, and data not shown). We speculate that this spacing could have a function in addition to the simple

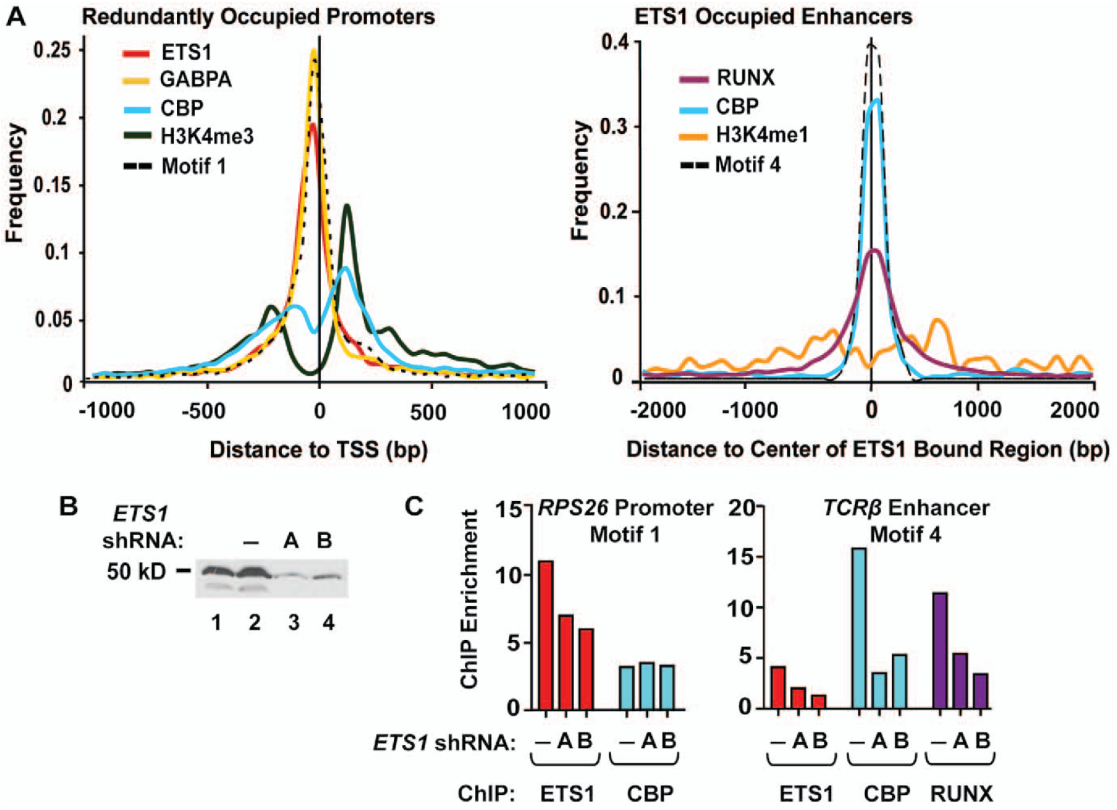


Figure 6. Distinct properties of promoters and enhancers occupied by ETS1. (A) Factor or histone modification positions were plotted as a class average across redundant promoters (left) or ETS1 occupied enhancers (right). At redundant promoters, the occupancy profiles of ETS1, GABPA, CBP, H3K4 tri-methyl, and Motif 1 were plotted from the center of each occupied region to the nearest RefSeq TSS. At ETS1-occupied enhancers, the occupancy profiles of RUNX1, CBP, H3K4 mono-methyl, and Motif 4 were plotted with respect to the center of the ETS1 bound region. For each factor, histone modification, or motif, a histogram of 30 bp bins was generated to represent the frequency of occupancy for each distance. The number of occurrences at each distance was normalized to the total number of regions with an occurrence of each factor, histone modification, or motif. A vertical line indicates the zero position of each chart (TSS or center of ETS1 bound region). (B) Protein immunoblot of Jurkat whole cell extracts with the ETS1 antibody. Lane 1, no shRNA; Lane 2, negative control shRNA targeting luciferase; Lanes 3 and 4, two independent shRNAs (A and B) targeting ETS1. The two bands apparent in Lanes 1 and 2 are consistent with the 51 and 42 kDa splicing isoforms of ETS1. (C) ETS1, CBP, and RUNX ChIP enrichment at the TCRb enhancer and the RPS26 promoter. The shRNAs were expressed in Jurkat T cells prior to ChIP, as indicated. Two independent biological replicates provided similar patterns, but different maximum levels of enrichment. A representative experiment is shown. Primer sequences used are provided in Table S3.
doi:10.1371/journal.pgen.1000778.g006

recruitment of ETS1 and RUNX. This may reflect a requirement for a specific conformation of ETS1 and RUNX for the transcriptional activation function of these enhancers. For example, because both ETS1 and RUNX can bind the co-activator CBP [15,16,45], and CBP occupies the same position as ETS1 and RUNX at enhancers (Figure 6A), cooperative CBP recruitment may require this particular configuration of ETS1 and RUNX.

A differential pattern of CBP enrichment at enhancers and promoters

This first picture of genome-wide occupancy of a transcription factor in combination with the co-factor CBP presented surprising diversity. In spite of the general picture that CBP occupancy was strongly correlated with ETS1 binding at both tissue specific enhancers and at active promoters in Jurkat T cells (Table 1), fine mapping uncovered a more complex picture (Figure 6A). The coincident binding observed at enhancers and the sensitivity of CBP occupancy at the TCR b enhancer to an ETS1 knockdown (Figure 6C) is consistent with direct recruitment of CBP by ETS1 and/or RUNX. These data are consistent with reports that CBP/p300 has a strong correlation with enhancers [21,30,46]. This not only supports the functionality of ETS1 bound distal enhancers, but also strongly demonstrates the role of DNA factors in CBP recruitment. In contrast, the lack of concordance of ETS1 and CBP binding events at promoters suggests that CBP is associated with other factors at these sites. Potential CBP recruitment mechanisms include interaction with general transcription factors [47] and enhancer-promoter looping [48,49]. Either of these mechanisms could contribute to the location of CBP at ETS1 bound promoters. We note, however, that the CpG island-containing promoters of housekeeping genes, at which we observe

redundant ETS occupancy, are thought to lack enhancers. Thus, we suggest that CBP is brought to these promoters by enhancer-independent interactions with the transcriptional machinery. One possibility is that ETS1 participates in recruitment, but maintenance at these constitutively active sites relies on cooperation with basal machinery or modified histones.

Binding site sequence variation guides diverse roles for transcription factors

Like many cellular proteins, transcription factors can have multiple roles that vary based on cell type and condition. Transcription factor function can also vary based on the context of other proteins present at each genomic locus. Here we show that the type of genes that are near ETS1 binding events, and the location of the co-activator CBP differ based on the sequence that recruits ETS1 to DNA. Thus, two different roles of ETS1 in T cells – a role at housekeeping promoters, and one at tissue specific enhancers – can be defined by distinct sequence motifs. The sequence variation for different functions of a transcription factor provides an explanation for the lack of a single binding sequence in many genome-wide occupancy studies. Our investigation provides a route to sort genome-wide binding data by the presence of such sequence motifs and other correlative data to define the distinct functions of a transcription factor.

Materials and Methods

ChIP

ChIP from Jurkat T cells was performed as described previously [18]. In brief, 56 10^7 cells were crosslinked with 1% formaldehyde and sheared chromatin extract was prepared. Dynabeads (Invitrogen) coupled to the appropriate secondary antibody were used to immunoprecipitate extracts treated with one of the following antibodies; polyclonal ETS1, sc-355; polyclonal CBP, A-22; (Santa Cruz Biotechnology), or monoclonal RUNX, a3.2.3.1. Crosslinks were reversed by heating and DNA was purified. Input controls were prepared in parallel, but with no immunoprecipitation step. qPCR analysis of ChIP DNA was performed as described previously [18]. In brief, the level of each region was determined by comparison to a standard curve of ChIP input DNA. Enrichments are a ratio of the level of the target region in each sample over the mean of the level of two negative control genomic regions.

Computational methods

The software used for ChIP-seq analysis is open source and available from the Useq project website (http://useq.sourceforge.net). Human annotation and sequence were obtained from the UCSC Genome Browser (March 2006, NCBI Build 36.1, HG18).

ChIP-seq analysis

ChIP and Input DNA was prepared for sequencing using Illumina's ChIP-seq kit. Each ChIP DNA sample was pooled from three independent replicates. 36 bp reads were generated using Illumina's Genome Analyzer II and standard pipeline software.

The following software from the Useq package [50] was used to identify regions enriched by ChIP compared to input control. ElandParser mapped sequence reads to the genome with an alignment score of. 13 (2 10 $\log_{10}$ (0.05)). FilterPointData was used to remove reads mapping to repeat regions included in the satellite repeat track from the UCSC genome browser (http://genome.ucsc.edu/). The number of non-repeat reads that mapped to the human genome for each sequencing sample was 6,683,411 for ETS1, 9,509,960 for RUNX, 8,525,775 for CBP, and 13,825,035 for input. PeakShiftFinder was used to measure the shift in the peak location between each DNA strand. ScanSeqs used a sliding window of 250 bp to score for enrichment across the genome and adjusted reads from opposite strands by 150 bp (ETS1, RUNX), or 125 bp (CBP), to remove the peak shift bias. EnrichedRegionMaker identified enriched regions. Significance was determined by calculating a binomial *P* value for each 250 bp window and controlled for multiple testing by calculating an empirical false discovery rate. The "Best Window" in each enriched region with an empirical false discovery rate of 0.01 were called as "bound regions" and had a median size of 250 bp.

Bound regions were overlapped using the IntersectRegions tool from Useq with no gap between regions except for overlaps reported in Table 1 in which a gap up to 300 bp was allowed. Enriched regions for ETS1, GABPA, and RUNX were screened for intersection with DNase I sensitive regions [19] before further analysis. This screening reduced the number of ETS1 bound regions from 19,420 to 14,824, the number of GABPA bound regions from 9214 to 7724, the number of CBP bound regions from 23,757 to 14,374, and the number of RUNX bound regions from 3632 to 1075. The P value for the overlap between ETS1 and DNase I sensitive sites was determined using IntersectRegions and comparing the ETS1 overlap to the overlap of 1000 iterations of a random regions of equivalent size derived from input point data. The P values for overlaps shown in Table 1 were derived by Fisher's exact test. The nearest Refseq TSS was determined using the FindNeighboringGenes tool. All ETS1 bound regions that intersect with DNase I sensitive regions are provided in Table S1. Table S1 also annotates the nearest gene, the presence or absence of Motif 1, 2 or 4 (Figure 3) and overlapping GABPA, RUNX, and CBP bound regions.

The ChIP-seq datasets and peak files are available for download from NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo), accession number GSE17954.

Protein purification and DNA binding assays

Protein purification and DNA binding assays were performed as described previously [18]. In brief, full-length human cDNAs of ETS1 (p51) and ELF1 were cloned into pet28a (Novagen) with an N-terminal 6x HIS tag, and expressed in bacteria. Proteins were isolated from inclusion bodies, resuspended in urea buffer (10 mM Tris (pH 7.9), 4 M urea, 500 mM NaCl, 15 mM immidazole), and bound to Ni-sepharose beads. After washing with urea buffer, protein was eluted with urea buffer with 750 mM immidazole and dialyzed overnight into 10 mM Tris (pH 7.9), 0.5 mM EDTA, 50 mM KCl, 1 mM DTT, and 10% glycerol. Protein concentration was determined by comparison to BSA standards on Coomassie brilliant blue stained SDS-PAGE gels. Three-fold serial dilutions of each protein were incubated with ^{32}P-labeled double stranded oligonucleotides (DNA concentration 6 $10^{2\ 11}$) for 30 min on ice and then run on a 6% polyacrylamide gel. Oligonucleotide probes had the following sequences (5-3): GGCCAAGCCGGAAGTGTGTGG-TAAACACTTT, GGCCAAGCCGGATGTGTGTGGTAAACA-CTTT,

GGCCAAGCAGGAAGTGTGTGGTAAACACTTT,

GGCCAAGCAGGATGTGTGTGGTAAACACTTT. K $_D$s were calculated by measuring the radioactivity in unbound bands by Phosphorimager (Molecular Dynamics) and using least squares curve fit analysis with fraction of DNA bound = 1/(1+ K_D/[Protein]).

Bioinformatic analysis by MEME, PATSER, and GOMINER

MEME (http://meme.sdsc.edu/meme4_1_1/cgi-bin/meme.cgi) was run with default settings except the minimum motif

length was set to 8 and the maximum motifl ength was set to 15. For each set of regions, the 250 with the highest log-transformed binomial P value for ETS1 were analyzed. Only the motif with the lowest E-value was reported.

The position weight matrixes for each motif used in the PATSER program (part of Regulatory Sequence Analysis Tools: http://rsat.ulb.ac.be/rsat/) to identify matches in each ETS1 bound region are listed in Table S2. PATSER score cutoffs used were Motif 1:9; Motif 2:12; Motif 4:10.3. Matches are listed in Table S1.

GoMiner (http://discover.nci.nih.gov/gominer/) was used to identify over-represented gene ontologies. All Refseq genes were used as the ''total'' and each subset of Refseq genes was used as the ''change'' file. Default settings were used except ''Evidence level 4'' and ''All/gene ontology'' were selected.

ETS1 knockdown

Two distinct small-hairpin RNAs (shRNAs) targeting ETS1, or a negative control shRNA targeting luciferase were cloned into pMK0.1p [51] and introduced to Jurkat T cells by MMLV based retroviruses. Jurkat T cells expressing the shRNA were selected by puromycin resistance. The sequences targeted were: luciferase, CTTACGCTGAGTACTTCGA; ETS1 A, AGGTGTAGACT-TCCAGAAG; ETS1 B, CTGATGTAAGGCAATTAAT.

Supporting Information

Figure S1 ETS1 bound regions have a high density within 500 bp of a TSS. The distance from the center of each of the 19,420 ETS1 bound regions to the nearest Refseq TSS was recorded. Distances were binned in 50 bp bins and frequencies plotted for bins from 0 to 5,000. Distances greater than 5,000 were discarded. 7,137 ETS1 bound regions had centers within 500 bp of a TSS.
Found at: doi:10.1371/journal.pgen.1000778.s001 (2.64 MB TIF)

Figure S2 Validation of ChIP-seq results by Q-PCR analysis of ChIP DNA. Enrichment was tested using primer sets specific for each region (Table S1). The level of each region was determined by comparison to a standard curve of ChIP input DNA. Enrichments are a ratio of the level of the target region in each sample over the mean of the level of two negative control genomic regions. The enrichments shown are the mean and standard error of the mean of two independent ChIP experiments. Regions were considered bound that had an mean enrichment of equal to or greater than 2. (A) 15 randomly selected ETS1 bound regions that overlapped with a DNase I sensitive site in CD**4** T cells (1–15), 8 randomly selected ETS1 bound regions that did not overlap with a DNase I sensitive site (16–23), and 9 randomly selected DNase I sensitive regions that were not scored as ETS1 bound (24–32) were tested for ETS1 ChIP enrichment in Jurkat T cells. Significantly more (13 of 15) ETS1 bound/DNase I sensitive regions were verified compared the other two categories to (0 of 8 and 0 of 9; P, 0.0001, Fisher's exact test). (B) 10 of 10 randomly selected CBP bound/DNase I sensitive regions were verified for CBP enrichment. (C) 8 of 8 randomly selected RUNX bound/DNase I sensitive regions were verified for RUNX enrichment.
Found at: doi:10.1371/journal.pgen.1000778.s002 (5.85 MB TIF)

Table S1 Properties of ETS1 bound regions. All chromosomal regions identified as ETS1 bound that also overlapped with DNase I sensitivity are listed. Regions that also overlapped with GABPA, RUNX, or CBP are indicated. Regions that contained Motif 1, 2, or 4 are indicated. The nearest RefSeq mRNA and gene, and the distance to the TSS is shown. The binomial P value and empirical false discovery rate reflect the significance of the ETS1 bound region.
Found at: doi:10.1371/journal.pgen.1000778.s003 (3.54 MB XLS)

Table S2 Position weight matrixes for Motif 1, 2, and 4.
Found at: doi:10.1371/journal.pgen.1000778.s004 (0.12 MB DOC)

Table S3 Oligonucleotide primers used in this study.
Found at: doi:10.1371/journal.pgen.1000778.s005 (0.07 MB DOC)

Acknowledgments

We thank University of Utah colleagues David Nix, for assistance with the Useq software package, and Brad Cairns, for critical reading of the manuscript.

Author Contributions

Conceived and designed the experiments: PCH KJC BJG. Performed the experiments: PCH KJC. Analyzed the data: PCH KJC RLP WEJ BJG. Contributed reagents/materials/analysis tools: NAS. Wrote the paper: PCH KJC NAS BJG.

References

1. Massie CE, Mills IG (2008) ChIPping away at gene regulation. EMBO Rep 9: 337–343.
2. Ji H, Vokes SA, Wong WH (2006) A comparative analysis of genome-wide chromatin immunoprecipitation data for mammalian transcription factors. Nucleic Acids Res 34: e146.
3. Rabinovich A, Jin VX, Rabinovich R, Xu X, Farnham PJ (2008) E2F in vivo binding specificity: comparison of consensus versus nonconsensus binding sites. Genome Res 18: 1763–1777.
4. Gaudet J, Mango SE (2002) Regulation of organogenesis by the Caenorhabditis elegans FoxA protein PHA-4. Science 295: 821–825.
5. Lam FH, Steger DJ, O'Shea EK (2008) Chromatin decouples promoter threshold from dynamic range. Nature 453: 246–250.
6. Leung TH, Hoffmann A, Baltimore D (2004) One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers. Cell 118: 453–464.
7. Meijsing SH, Pufall MA, So AY, Bates DL, Chen L, et al. (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. Science 324: 407–410.
8. Galang CK, Muller WJ, Foos G, Oshima RG, Hauser CA (2004) Changes in the expression of many Ets family transcription factors and of potential target genes in normal mammary tissue and tumors. J Biol Chem 279: 11281–11292.
9. Hollenhorst PC, Jones DA, Graves BJ (2004) Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. Nucleic Acids Res 32: 5693–5702.
10. Sharrocks AD (2001) The ETS-domain transcription factor family. Nat Rev Mol Cell Biol 2: 827–837.
11. Oikawa T, Yamada T (2003) Molecular biology of the Ets family of transcription factors. Gene 303: 11–34.
12. De Val S, Chi NC, Meadows SM, Minovitsky S, Anderson JP, et al. (2008) Combinatorial regulation of endothelial gene expression by ets and forkhead transcription factors. Cell 135: 1053–1064.
13. Chen C, Ouyang W, Grigura V, Zhou Q, Carnes K, et al. (2005) ERM is required for transcriptional control of the spermatogonial stem cell niche. Nature 436: 1030–1034.
14. Muthusamy N, Barton K, Leiden JM (1995) Defective activation and survival of T cells lacking the Ets-1 transcription factor. Nature 377: 639–642.
15. Yang C, Shapiro LH, Rivera M, Kumar A, Brindle PK (1998) A role for CREB binding protein and p300 transcriptional coactivators in Ets-1 transactivation functions. Mol Cell Biol 18: 2218–2229.
16. Jayaraman G, Srinivas R, Duggan C, Ferreira E, Swaminathan S, et al. (1999) p300/cAMP-responsive element-binding protein interactions with ets-1 and ets-2 in the transcriptional activation of the human stromelysin promoter. J Biol Chem 274: 17342–17352.
17. Foulds CE, Nelson ML, Blaszczak A, Gr aves BJ (2004) MAPK phosphorylation activates Ets-1 and Ets-2 by CBP/p300 recruitment. Mol Cell Biol 24: 10954–10964.
18. Hollenhorst PC, Shah AA, Hopkins C, Graves BJ (2007) Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. Genes Dev 21: 1882–1894.
19. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, et al. (2008) High-resolution mapping and characterization of open chromatin across the genome. Cell 132: 311–322.

20. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, et al. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129: 823–837.
21. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, et al. (2007) Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet 39: 311–318.
22. Robertson AG, Bilenky M, Tam A, Zhao Y, Zeng T, et al. (2008) Genome-wide relationship between histone H3 lysine 4 mono- and tri-methylation and transcription factor binding. Genome Res 18: 1906–1917.
23. Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, et al. (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods 5: 829–834.
24. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. Proceedings of the Second International Conference on Intellegent Systems for Molecular Biology. Menlo Park, CA: AAAI Press. pp 28–36.
25. Wang CY, Bassuk AG, Boise LH, Thompson CB, Bravo R, et al. (1994) Activation of the granulocyte-macrophage colony-stimulating factor promoter in T cells requires cooperative binding of Elf-1 and AP-1 transcription factors. Mol Cell Biol 14: 1153–1159.
26. Goetz TL, Gu TL, Speck NA, Graves BJ (2000) Auto-inhibition of Ets-1 is counteracted by DNA binding cooperativity with core-binding factor alpha2. Mol Cell Biol 20: 81–90.
27. Giese K, Kingsley C, Kirshner JR, Grosschedl R (1995) Assembly and function of a TCR a enhancer complex is dependent on LEF-1 induced DNA bending and multiple protein-protein interactions. Genes Dev 9: 995–1008.
28. Hertz GZ, Stormo GD (1999) Identifying DNA and protein patterns with statistically significant alignments of multiple sequences. Bioinformatics 15: 563–577.
29. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, et al. (2003) GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol 4: R28.
30. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457: 854–858.
31. Jiang C, Pugh BF (2009) Nucleosome positioning and gene regulation: advances through genomics. Nat Rev Genet 10: 161–172.
32. Bosselut R, Levin J, Adjadj E, Ghysdael J (1993) A single amino-acid substitution in the ETS domain alters core DNA binding specificity of Ets-1 to that of the related transcription factors Elf-1 and E74. Nucleic Acids Res 21: 5184–5191.
33. Yokomori N, Kobayashi R, Moore R, Sueyoshi T, Negishi M (1995) A DNA methylation site in the male-specific P450 (Cyp 2d-9) promoter and binding of the heteromeric transcription factor GABP. Mol Cell Biol 15: 5355–5362.
34. Gaston K, Fried M (1995) CpG methylation has differential effects on the binding of YY1 and ETS proteins to the bi-directional promoter of the Surf-1 and Surf-2 genes. Nucleic Acids Res 23: 901–909.
35. Bird A (2002) DNA methylation patterns and epigenetic memory. Genes Dev 16: 6–21.
36. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, et al. (2008) Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature 454: 766–770.
37. Rozenberg JM, Shlyakhtenko A, Glass K, Rishi V, Myakishev MV, et al. (2008) All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. BMC Genomics 9: 67.
38. Bories J-C, Willerford DM, Grevin D, Davidson L, Camus A, et al. (1995) Increased T cell apoptosis and terminal B-cell differentiation induced by inactivation of theEts-1 proto-oncogene. Nature 377: 635–638.
39. Barton K, Muthusamy N, Fischer C, Ting C-N, Walunas T, et al. (1998) The Ets-1 transcription factor is required for the development of natural killer cells in mice. Immunity 9: 555–563.
40. Guo Y, Maillard I, Chakraborti S, Rothenberg EV, Speck NA (2008) Core binding factors are necessary for natural killer cell development and cooperate with Notch signaling during T cell–specification. Blood 112: 480–492.
41. Egawa T, Eberl G, Taniuchi I, Benlagha K, Geissmann F, et al. (2005) Genetic evidence supporting selection of the Valpha14i NKT cell lineage from double-positive thymocyte precursors. Immunity 22: 705–716.
42. Ohno S, Sato T, Kohu K, Takeda K, Okumura K, et al. (2008) Runx proteins are involved in regulation of CD122, Ly49 family and IFN-gamma expression during NK cell differentiation. Int Immunol 20: 71–79.
43. Wotton D, Ghysdael J, Wang S, Speck NA, Owen MJ (1994) Cooperative binding of Ets-1 and Core Binding Factor to DNA. Mol Cell Biol 14: 840–850.
44. Sun W, Graves BJ, Speck NA (1995) Transactivation of the Moloney murine leukemia virus and T cell receptorb-chain enhancers bycbfand etsrequires intact binding sites for both proteins. J Virol 69: 4941–4949.
45. Kitabayashi I, Yokoyama A, Shimizu K, Ohki M (1998) Interaction and functional cooperation of the leukemia-associated factors AML1 and p300 in myeloid cell differentiation. EMBO J 17: 2994–3004.
46. Wang Z, Zang C, Cui K, Schones DE, Barski A, et al. (2009) Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. Cell 138: 1019–1031.
47. Black JC, Choi JE, Lombardo SR, Carey M (2006) A mechanism for coordinating chromatin modification and preinitiation complex assembly. Mol Cell 23: 809–818.
48. Resendes KK, Rosmarin AG (2006) GA-binding protein and p300 are essential components of a retinoic acid-induced enhanceosome in myeloid cells. Mol Cell Biol 26: 3060–3070.
49. Teferedegne B, Green MR, Guo Z, Boss JM (2006) Mechanism of action of a distal NF-kappaB-dependent enhancer. Mol Cell Biol 26: 5759–5770.
50. Nix DA, Courdy SJ, Boucher KM (2008) Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. BMC Bioinformatics 9: 523.
51. Masutomi K, Yu EY, Khurts S, Ben-Porath I, Currier JL, et al. (2003) Telomerase maintains telomere structure in normal human cells. Cell 114: 241–253.
52. Su AI, Cooke MP, Ching KA, Hakak Y, Walker JR, et al. (2002) Large-scale analysis of the human and mouse transcriptomes. Proc Natl Acad Sci U S A 99: 4465–4470.

CHAPTER 3

DISRUPTION OF ETS1 OCCUPANCY REVEALS DISTINCT BEHAVIORS OF ETS FACTOR TARGET GENES

Abstract

Understanding how transcription factor families mediate distinct functions despite highly conserved DNA binding site preferences is a central question in gene regulation. With a focus on ETS1, we previously established a model of ETS factor function involving two distinct modes of recruitment: redundant ETS recruitment using a high-affinity consensus site, and specific ETS1 recruitment using a low-affinity degenerate site. Here, we disrupt ETS1 binding *in vivo* to determine the molecular basis of redundancy and refine our model of ETS factor function. At redundant sites, increased GABPA occupancy compensated for disruption of ETS1 occupancy and global expression analysis verified constitutive transcription of nearby genes. At ETS1 specific sites, GABPA did not compensate for disruption of ETS1 occupancy, and neighboring genes were unexpectedly upregulated. These findings build on the previous model of ETS family function in two ways: 1) they suggest that redundancy represents a time sharing mechanism where ETS factors take turns sampling an individual binding site, and 2) they highlight that specificity within the ETS family is likely a continuum rather than a strict dichotomy. We hypothesize that ETS2 may play a role in the upregulation of ETS1 specific target genes, and future studies will address the continuum of specificity for highly conserved family members like ETS1 and ETS2.

Introduction

Metazoan genomes are characterized by groups of related genes that are derived from gene duplications during evolution. The diversification of function of these gene families enhanced the capacity of evolving metazoans to achieve new tissue types and other complexities. Yet, within these gene groups functional domains often retain highly conserved properties. Functional domains, such as DNA binding domains, are subject to selective conservation during evolution, and transcription factor families often exhibit a conserved preference for DNA binding sequences [1]. Transcription factor families vary widely in size, from the extremely large nuclear receptor group to the four-member Sp group, but nearly all families have been implicated in mediating a range of individual functions [2,3,4,5,6]. Thus, understanding how highly related transcription factors perform distinct functions is a central issue in the overall understanding gene regulation.

The ETS family of transcription factors provides a model system to address this conundrum of specificity. There are 28 *ets* genes in humans, all containing a highly conserved DNA binding domain, the ETS domain, which preferentially binds a core DNA sequence, 5'-GGA(A/T)-3'. The family is widely represented across metazoans; there are 8 *ets* genes in *Drosophila melanogaster* [7], 10 in *Caenorhabditis elegans* [8], and 28 in *Homo sapiens* [6]. Mouse knockout models show distinct defects in mice with individual ETS factor deletions, suggesting unique functions for different members of the ETS family. Mouse phenotypes highlight the diverse biology of the ETS family, and range

from male infertility upon ETV4/5 knockout [9,10] to hematopoietic cell defects upon ETS1 knockout [11,12]. This specificity is not driven by cell type restricted expression because a given cell type expresses multiple ETS factors, representing a diverse network of ETS factors in an individual cell [13,14]. Thus, it is clear that molecular and genomic mechanisms exist to determine specific functions within the context of the ETS network, and *in vitro* and *in vivo* studies have begun to shed light on these mechanisms.

In vitro studies that select high-affinity binding sites have shown that binding site preferences across the ETS family are nearly identical [15,16,17,18,19,20]. More recently, high-throughput protein binding microarrays have been used to assess the DNA binding preferences of 27 human ETS factors [21]. The factors were classified into four distinct clades, and like previous experiments, this study found remarkable concordance of the sites recognized by the family. ETS factors in Clade I, a subgroup that contains over half of all ETS factors, showed no distinguishing preferences for the core DNA binding site. Slight differences in preference were observed in the three ETS factors in Clade III, PU.1, SPIB, and SPIC, which preferred a 5'-GGAA-3' binding site, and the sole ETS factors in Clade IV, SPDEF, which preferred a 5'-GGAT-3' binding site. These differences were subtle, however, and it is unlikely that preferences of the core DNA recognition sequence provide adequate specificity to drive the diverse biological functions of the ETS family.

While degenerate residues flanking the 5'-GGA-3' of the ETS binding site can make a site thermodynamically unfavorable, divergence from high-affinity sites could provide greater flexibility to discriminate between ETS factors. Indeed, *in vitro* studies on individual ETS factors indicate that specificity is mediated by low-affinity binding sites. For example, structural studies have shown that specific contacts between the PU.1 ETS domain and degenerate residues flanking the core ETS sequence drive PU.1 specificity at a low-affinity binding site [22,23]. Cell based reporter experiments suggest that a low-affinity binding site in the promoter of the IL2 gene, 5'-ATATAGGAAGTG-3', mediates ELF1 specific occupancy and transcriptional regulation [16]. Mutation of this site to a high-affinity ETS site decreases transcription, possibly through interference by other ETS factors like FLI1.

In vivo experiments, including the determination of genome wide occupancy profiles for a handful of ETS factors, are beginning to reveal distinct biological roles for high- and low- affinity binding sites. At high-affinity binding sites, a striking degree of ETS factor redundancy has been observed, despite the fact that diverse ETS factors have been considered [21,24,25,26]. Redundant occupancy of ETS factors is most frequently observed between 20 bp and 40 bp upstream of transcription start sites (TSS) of genes associated with 'house-keeping' (HK) functions [24,25]. Fine mapping of the occupancy of ETS1 and GABPA reveals strong co-localization of the factors, occurring at a motif closely matching the *in vitro* derived ETS consensus site [25].

In contrast to the redundant function of high-affinity sites, low-affinity sites are associated with the occupancy of a specific ETS factor and can involve heterotypic interactions such as the interaction between ETS1 and RUNX1 [24,25] or between ELK1 and SRF [26]. Specific binding events frequently occur distal to the nearest TSS (>500 bp), and in the case of ETS1, neighboring genes are associated with cell type specific hematopoietic function. These results, as well as *in vitro* experiments, contribute to the working model of ETS family function: ETS factors redundantly regulate constitutive HK gene expression through high-affinity, promoter proximal binding sites, and individual ETS factors regulate cell-type specific expression through low-affinity, promoter distal binding sites.

This model sets a frame work for ETS factor function, but several key features of the model remain unclear. While analysis of individual HK genes has suggested active transcription, genome-wide confirmation that redundant ETS targets are actively transcribed has not been established. The promoters of HK genes are correlated with identifiers of active promoters and constitutive expression, such as DNaseI hypersensitivity, H3K4me3, and high promoter PolII occupancy across diverse cell types [27,28,29,30], but comparative probe-based gene expression techniques (e.g., microarray) are not sufficient to quantitatively establish expression of genes that are constitutively expressed across all cell types. Thus, while the model of ETS factor function suggested that active

transcription is mediated from redundant binding sites, this hypothesis had not been confirmed.

Additionally, while it is established that redundant ETS factor occupancy occurs at the same binding site [25], the molecular mechanism of how multiple ETS factors 'share' a site was unclear. We hypothesized that the observed redundancy is likely a result of one of two mechanisms: 1) different ETS factors occupy the site in different cells and the observed redundancy reflects average occupancy in a heterogeneous population; or 2) different ETS factors time share a single site in a dynamic equilibrium within a single cell. If the latter model is correct, disruption of one ETS factor would shift the equilibrium of ETS factor occupancy. Thus, these two models could be distinguished by disrupting an ETS factor *in vivo* and observing the behavior of other ETS factors. To date, however, traditional RNAi has not been successful in distinguishing this model because most ETS factors contribute to the expression of other ETS factors, resulting in off target effects. In this study, we postulate that phosphorylation dependant inhibition of ETS1 DNA binding, which has been studied extensively *in vitro*, may serve as a tool to overcome the limitations of RNAi and allow rapid, *in vivo* ETS network perturbation. Activated calcium/calmodulin dependant kinase II (CaMKII) phosphorylates ETS1 at multiple serines in a flexible, serine rich region (SRR) near the ETS domain [31], resulting in a 10-fold reduction of the DNA binding-affinity of ETS1 [32,33,34]. While the phosphorylation of ETS1 in cells has been characterized in mouse cell lines [31] and *in vivo* calcium signaling is correlated

with reduced expression of an ETS1-driven reporter construct [35], the global impact of ETS1 phosphorylation *in vivo* was not understood.

Here we seek to develop calcium signaling as a biological tool for perturbing the ETS network *in vivo*, both to validate the functional relevance of ETS1 phosphorylation *in vivo* and to distinguish between the two models of redundancy. Specifically, we consider how ETS1 disruption affects the equilibrium of ETS factor occupancy and the expression of previously identified ETS target genes. Upon disruption of ETS1 occupancy, we observed enhanced occupancy of GABPA at redundant sites but not at ETS1 specific sites. Using RNASeq, we determined that the expression of ETS-redundant target genes was unaffected by the shift in ETS factor equilibrium, suggesting that GABPA binding is compensatory and functionally redundant. At ETS1-specific target genes, disruption of ETS1 binding was unexpectedly correlated with transcriptional upregulation, and future studies will consider the basis of this increased transcription. This finding indicates that site preference in the ETS family is a continuum of specificity rather than a dichotomy of redundant and specific sites.

Results

Perturbing the ETS network in T cells

The pharmacological stimulation of Jurkat T cells was used as a model to induce calcium-dependent phosphorylation of ETS1. To establish the model, we confirmed previous *in vivo* studies on calcium signaling [31,36] and established a time course of ETS1 phosphorylation in our system. Jurkat cells were activated

with phorbol-12-myristate-13-acetate (PMA) and ionomycin calcium ionophore, and harvested at the following time points: 2 min, 10 min, 30 min, 1 hr, and 3 hr. Western blot analysis was performed on whole cell lysates of the sampled populations (Figure 3.1). Long-exposure analysis of total ETS1 revealed two major ETS1 isoforms expressed in Jurkat cells (exposure not shown), previously identified and referred to as p42 and p51 based on electrophoretic mobility [37,38]. After 2 min of activation, we observed a mobility shift of p51, but not p42. The shorter ETS1 isoform lacks exon VII, which contains the SRR, indicating that the shift in p51 was likely due to ETS1-SRR phosphorylation. Western blot with a phosphospecific antibody for ETS1-S251**P**, one of the SRR phosphorylation events, verified that the shifted species is ETS1-SRR phosphorylated. Note the complete absence of ETS1 calcium-dependent phosphomark in the resting state. After 30 min of activation, the p51 species was entirely shifted, indicating that all or nearly all of the ETS1 in the cell was SRR-phosphorylated. After 1 hr of activation, the phosphomark declines and was not detectable by 3 hr. Between 5 hr and 8 hr postactivation, ETS1 levels declined, a phenomena previously reported by [39].

ETS1-SRR phosphorylation reduces global

ETS1 DNA binding *in vivo*

To determine if ETS1-SRR phosphorylation reduces DNA binding *in vivo*, we performed ETS1 ChIPSeq on cells activated for 30 min and compared the data to a previous ETS1 ChIPSeq experiment in resting cells [25] (Figure 3.2a).

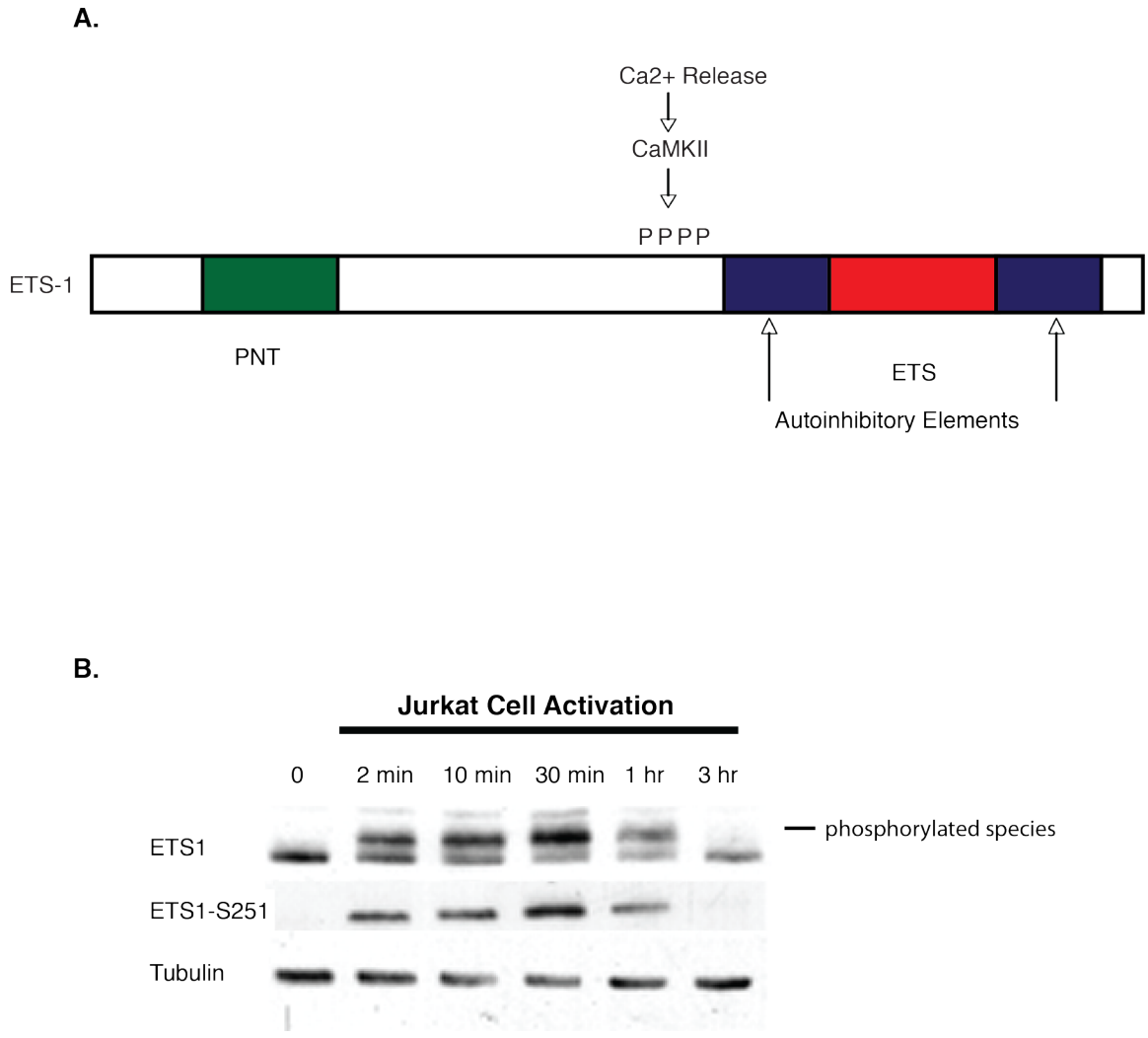
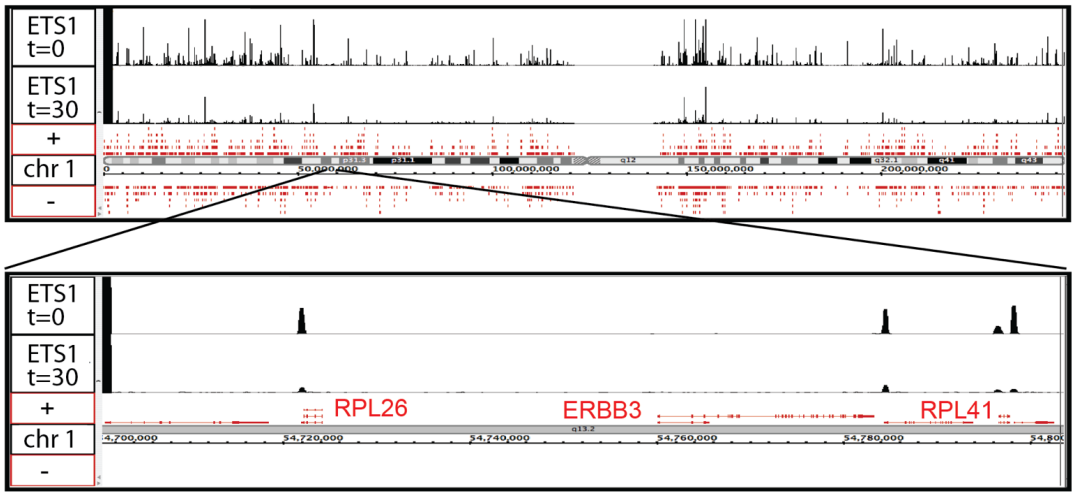


Figure 3.1 Calcium signaling to ETS1 (A) Calcium-dependant signaling impinges on ETS1, resulting in multiple discrete phosphorylation events in a serine rich region (SRR) N-terminal to the ETS domain [31]. (B) A time course of activation followed by protein immunoblot of Jurkat whole cell extract reveals the rapid modification of ETS1. Analysis with ETS1 antibody reveals an electrophoretic mobility shift after 2 min of stimulation and peaking at 30 min of phosphorylation. Analysis with an antibody specific for phospho-ETS1S251 confirms that the observed shift is correlated with to SRR-phosphorylation.

Using an empirical false discovery rate of < 0.01, there were 20,439 binding events in the activated data set and 19,420 in the resting set. Furthermore, 68% of the bound regions in the activated set were included in the resting set. Of the nonoverlaping binding events in the activated data set, 87% are not DNase I hyper-sensitive [27] and, therefore, may represent false positives (Figure 3.2b). In spite of the high number of identified binding events, the average binomial P value in the activated set was 77.5 compared to 101.0 in the resting set, indicating that ETS1 occupancy was decreased upon activation. Qualitative analysis of the resting and activated ETS1 ChIPSeq data sets confirms this quantitative analysis. In sum, these data indicate a global reduction in ETS1 occupancy upon activation.

These findings were confirmed and quantified by quantitative PCR (qPCR) on ETS1-specific and ETS-redundant bound regions (Figure 3.3). ETS1 occupancy was observed to dramatically decline at 3/3 ETS-redundant promoters and at 3/4 ETS1-specific enhancers. We calculated that the average decrease in ETS1 occupancy at promoters is 72$\pm$ 21% while for enhancers it is 46$\pm$23%. While this difference may be significant, the range of qPCR sensitivity at high occupancy sites (promoters) and low occupancy sites (enhancers) may not be comparable. These data confirm that ETS1 occupancy is disrupted upon calcium signaling *in vivo*.

A.



B.

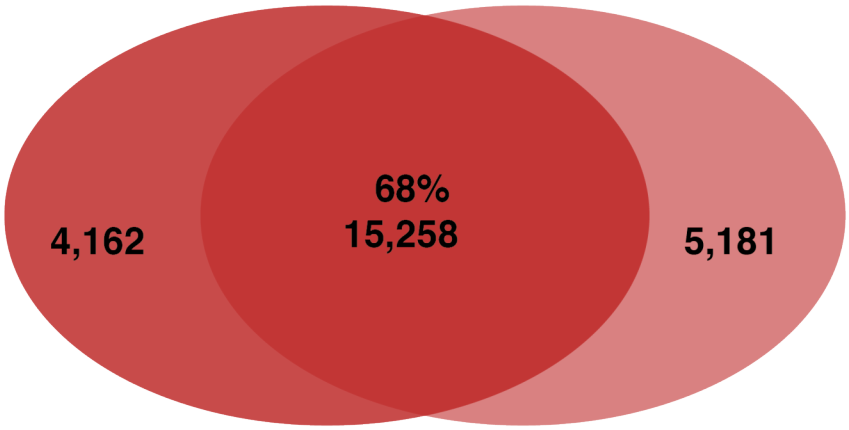
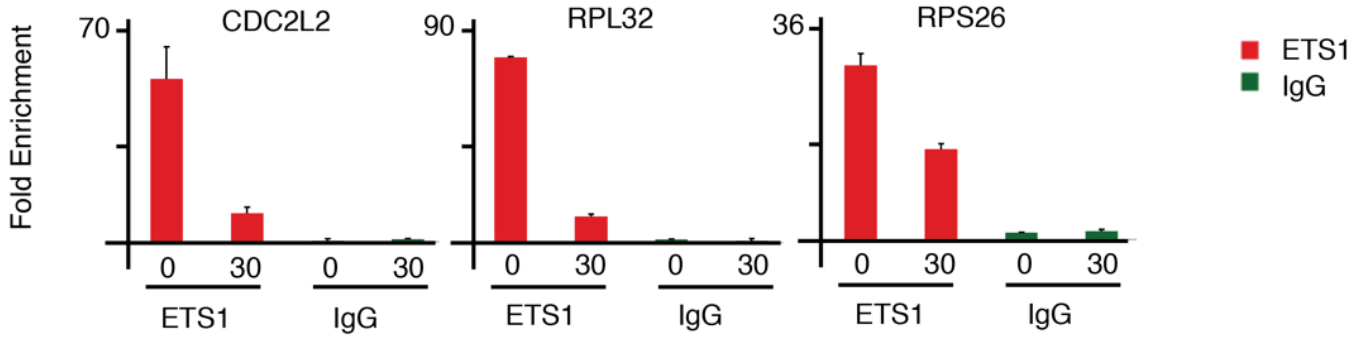


Figure 3.2 Signaling disrupts ETS1 occupancy genome wide (A) Log transformed P values of ETS1 occupancy in resting [25] and 30 min activated cells. Mapped sequence reads (NCBI build 36.1) are visualized using the integrated genome browser (http://igb.bioviz.org/). (B) With an empirical false discovery rate of < 0.01, 19,420 and 20,439 ETS1 binding events were identified in the resting and activated data sets, respectively, and 68% of the activated binding sites were contained within the resting set. The average binomial P-value associated with occupancy is 77.5 for the activated set, and 101.0 for the resting set.

ETS Factor Redundant Sites: average ETS1 decrease of 72% ± 21% (n=3)



ETS1 Specific Sites: average ETS1 decrease of 46% ± 23% (n=4)

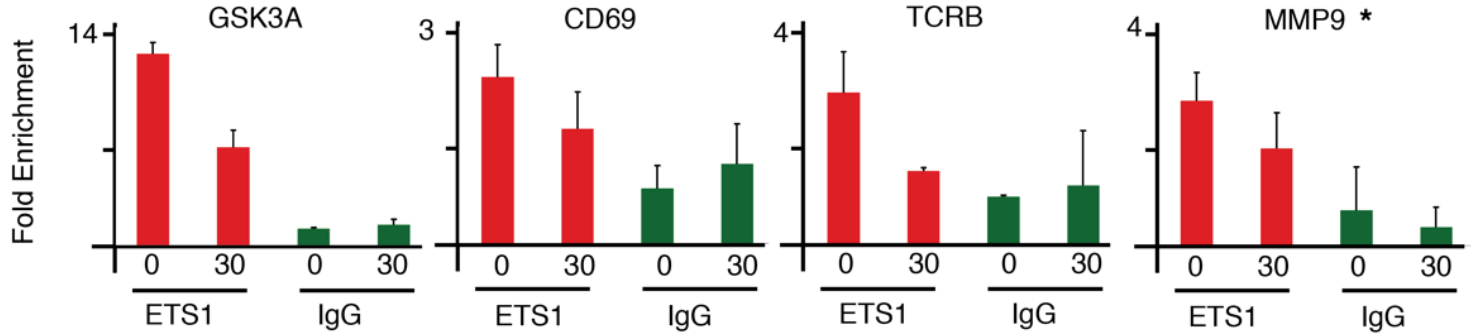


Figure 3.3 Signaling disrupts ETS1 occupancy on individual targets ChIPSeq data was confirmed by performing qPCR on direct target genes. At 3/3 redundant promoters and 3/4 ETS1 specific genes, ETS1 occupancy was decreased > 50% upon signaling. The average decrease in ETS1 occupancy at promoters was 72± 21% while at enhancers it was 46±23%.

* ETS1 decrease < 50% upon signaling.

Disruption of ETS1 tests the time-share model of redundancy

We have proposed that redundant ETS factor occupancy could reflect time-sharing of an individual ETS site in a cell or the average occupancy of each factor across the cell population. Upon ETS1 disruption, compensation by other ETS factors would support a model of time-sharing on an individual site. Previous studies showed that GABPA is highly expressed in Jurkat T cells [14] and redundantly occupies housekeeping promoter sites with ETS1 [24,25], making the factor a good candidate for examining compensation by ETS factors. As described below, these two ETS factors retain a presence in Jurkat cells after activation.

We performed ChIP and qPCR to query ETS1 and GABPA occupancy in resting and activated cells at 30 min, representing the peak of ETS1-SRR phosphorylation. At 3/3 redundant promoter sites, decreased ETS1 occupancy after 30 min of activation was correlated with increased GABPA occupancy (Figure 3.4). Further, these results were not recapitulated on ETS1 specific binding sites (0/4). This experiment provides an example of ETS factor time sharing on an individual site, and validates our previous classification of specific and redudant ETS1 binding sites.

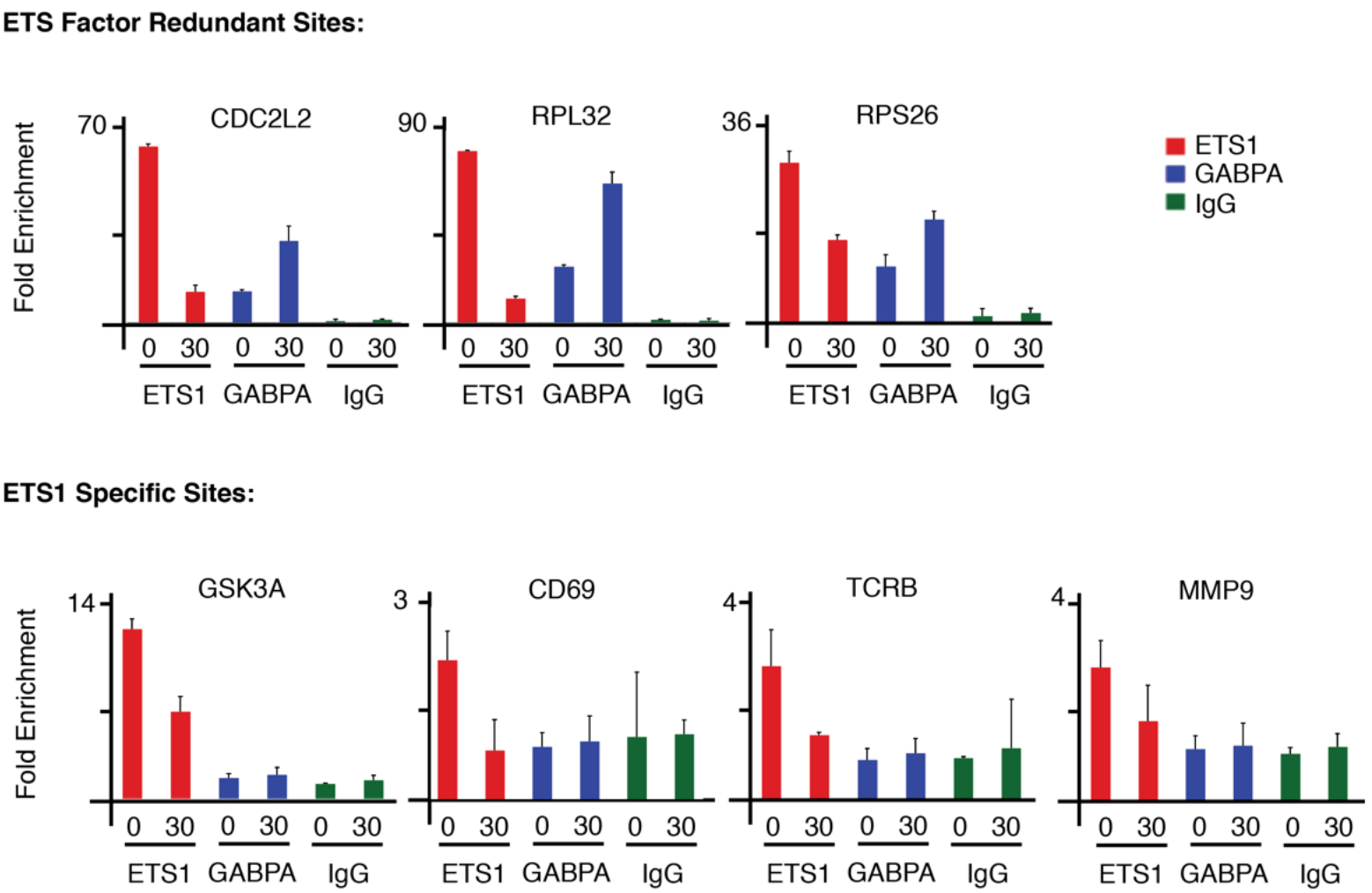


Figure 3.4 GABPA occupancy is increased as ETS1 occupancy is decreased At 3/3 redundant promoters and 0/4 ETS1 specific enhancers, decreased ETS1 occupancy is correlated with increased GABPA occupancy. At redundant sites, GABPA occupancy was increased an average of 2.3-fold upon signaling, while at specific sites GABPA only increased 1.06-fold.

The expression of genes near redundant ETS sites

are not changed by ETS network perturbation

ETS1 and GABPA redundantly time share high-affinity ETS sites in the promoters of HK genes, but do they redundantly regulate transcription? We hypothesize that redundancy in the ETS family establishes constitutive expression of HK genes [24,25]. To test whether ETS-factor target genes are expressed in resting Jurkat T cells, we employed an RNASeq approach, which enables absolute values of steady-state RNA levels to be quantified as reads per kilobase of exon model per million mapped reads, or RPKM, reflecting the number of sequence reads for a transcript normalized to its length [40].

RNASeq was performed on poly-A selected RNA from resting Jurkat T cells. We mapped the expression data to annotated RefSeq mRNA transcripts and plotted a histogram of the RPKM associated with each transcript to determine a threshold for defining expression (Figure 3.5a). Using the selected threshold of RPKM ≥ 5, we identified that 78% (2283/2927) of redundant ETS-target genes were expressed, compared to only 36% (7984/21989) of the reference set, which contained all RefSeq genes ($P < .0001$). We also found that increased redundancy is correlated with increased expression. For example, genes neighboring sites occupied by ETS1, ELF1, and GABPA had an average RPKM value of 49.1 while genes neighboring only one ETS factor had a lower average RPKM (Figure 3.5b). Genes neighboring sites bound by two ETS factors

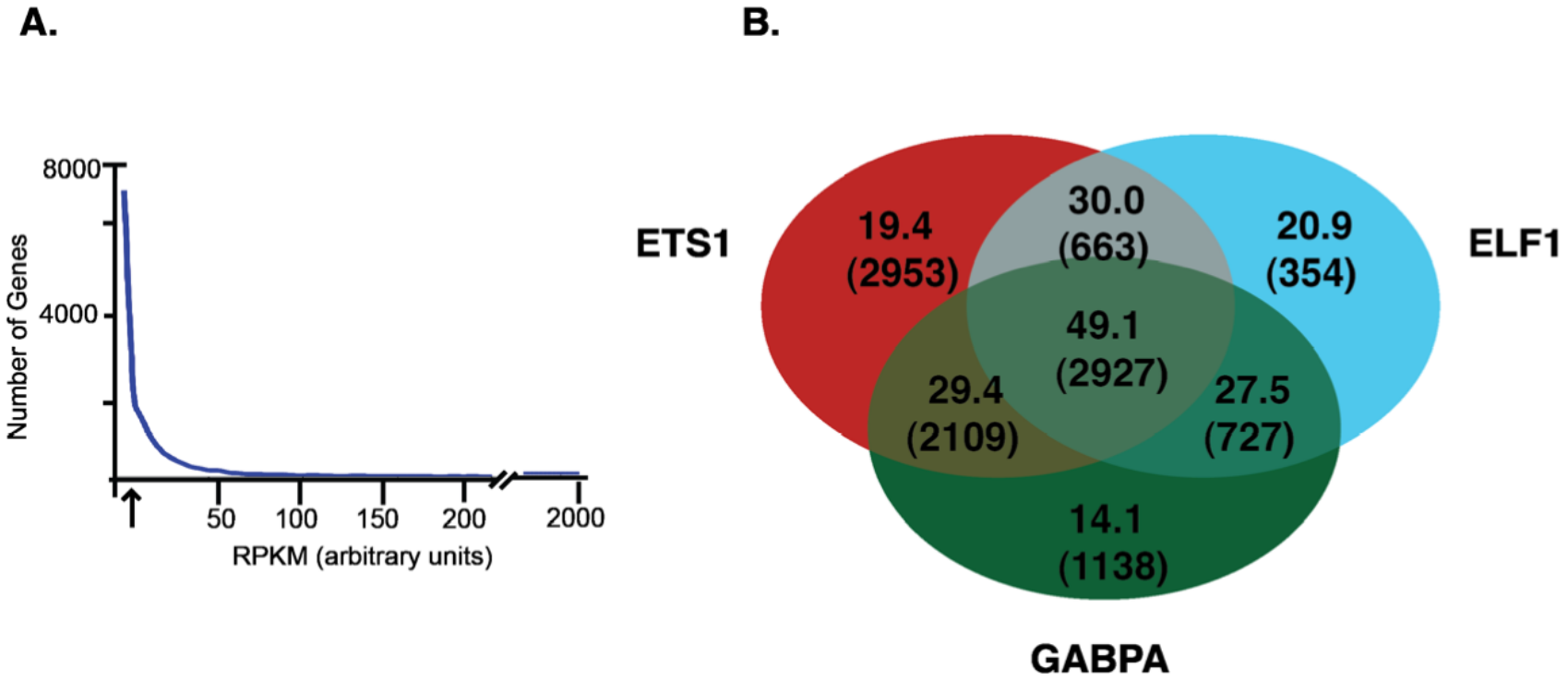


Figure 3.5 Redundant ETS factor occupancy is correlated with positive transcription (A) The absolute expression (defined as Reads Per Kilobase of exon model per Million mapped reads, or RPKM) of annotated RefSeq mRNA transcripts was determined by mapping sequenced RNA from resting Jurkat T cells to the genome (NCBI build 36.1) [65]. RPKM values were placed in 1-unit bins and plotted as a histogram. The majority of transcripts (64%) are expressed at levels less than 5 RPKM, which was selected as a semi-arbitrary threshold to define active transcription. (B) Correlation of ETS factor occupancy with RPKM shows that genes neighboring sites occupied by ETS1, ELF1, and GABPA [21, 25, 41] have an average RPKM value of 49.1 (2,927 regions), while genes neighboring sites occupied by only one or two ETS factor(s) have lower average RPKM values. For example, genes near ETS1 and ELF1 occupied regions have an average RPKM of 30.0 (663 regions), while those near ELF1-only occupied regions have an average RPKM of 20.9 (354 regions).

had an intermediate value of expression. Thus, we conclude that genes neighboring ETS redundant binding events are actively transcribed.

Next, we sought to determine if the expression of ETS redundant target genes are affected by disrupted ETS1 occupancy. Jurkat T cells were activated for a 12 hr time course and RNASeq was performed on samples collected at the following time points: 30 min, 1 hr, 3 hr, 5 hr, 8 hr, and 12 hr. Again, the expression data was mapped to annotated RefSeq transcripts. We generated a class average expression profile of 1) transcripts proximal to ETS redundant sites and 2) a control gene set representing all RefSeq mRNAs (Figure 3.6a). The expression profile represents the fold-change in RPKM relative to the resting state for the gene sets. Over the 12-hr time course, the redundant ETS target gene set and the reference gene set exhibit the same unchanged expression profile. We conclude that the expression of ETS redundant target genes does not change upon ETS1 disruption, and we speculate that ETS1 and GABPA, and perhaps other ETS factors, are functionally interchangeable at these sites.

ETS1 specific target genes are upregulated upon ETS network perturbation

In resting Jurkat T cells, RNASeq data indicated that 66% (644/975) of specific ETS1-target genes were expressed above the RPKM ≥ 5 threshold. As in the case of the ETS redundant gene set, this represents an enrichment over the 36% expression in the reference set ($P < .0001$). The fact that fewer ETS1 specific targets were expressed than ETS redundant targets could be an artifact

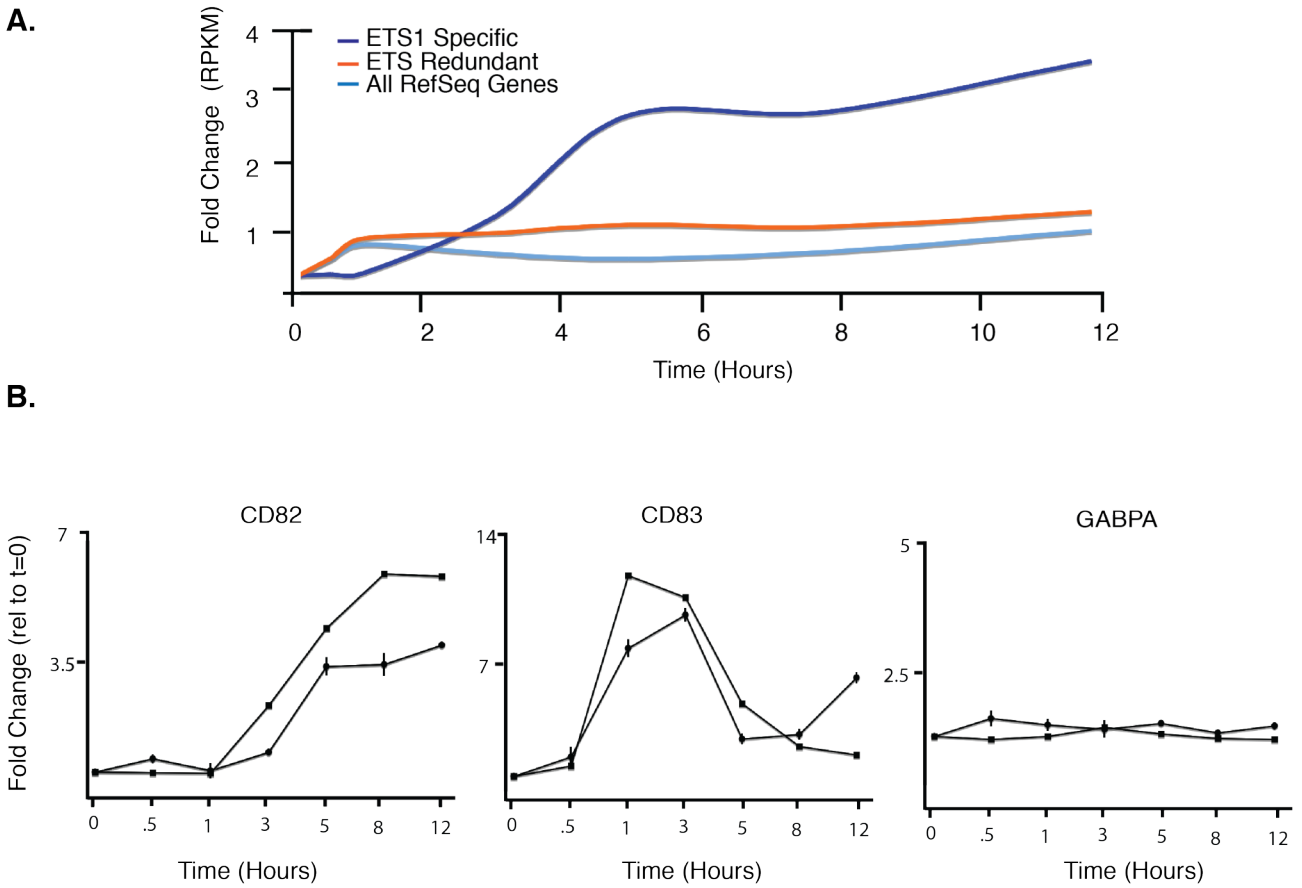


Figure 3.6 Global expression analysis reveals distinct behaviors of ETS targets (A) Jurkat T cells were activated for a 12 hr time course and RNASeq was performed on samples collected at the following time points: 30 min, 1 hr, 3 hr, 5 hr, 8 hr, and 12 hr. Data was mapped to annotated RefSeq transcripts, and fold change relative to resting cells was determined for genes near redundant ETS binding sites and ETS1 specific binding sites, with All RefSeq genes serving as a control. Genes near ETS redundant binding sites were not observed to change over the time course, but genes near ETS1 specific binding sites were found to be significantly upregulated by 5 hr of activation. (B) A subset of 10 genes were analyzed by qPCR to confirm RNASeq findings, and 10/10 showed the similar profiles to the RNASeq data. Three representative genes are shown.

of incorrect gene assignment to bound region, which was made arbitrarily based on proximity.

Because we did not observe GABPA compensation at ETS1 specific sites, we predicted that that the expression of ETS1 specific targets would decrease over the time course of expression. As described above, we generated a profile for the change in expression the ETS1 specific target gene set over the RNASeq time course. Surprisingly, the class average of RNA expression of ETS1 specific target genes increased between 3 and 5 hr of activation (Figure 3.6a), a time point at which ETS1 protein is decreased. Between 3 hr and 8 hr of activation, 81% (790/957; $P < .0001$) of genes in this class were upregulated more than 2-fold, and 42% (409/790; $P = 0.0018$) were upregulated more than 3-fold. The RNASeq findings were confirmed for 10/10 genes of genes by RT-qPCR (subset shown Figure 3.6b).

There are many possible explanations for why ETS1 specific target genes are upregulated at time points depleted for ETS1 protein. Here, we explore only one of several possible scenarios: that an ETS factor other than ETS1 or GABPA mediates this transcriptional upregulation.

ETS2 may control of a subset of nonredundant genes

The genome wide binding profiles of only a handful of ETS factors have been characterized [21,25,41,42]. We determined the profile of expression of all *ets* genes across the RNASeq time course to identify candidate factors for mediating the observed upregulation. Comparison of the resting state RNASeq

data to previously determined mRNA copy number for each *ets* gene [14] found good correlation between the two methods (Figure 3.7), with 13 genes expressed at levels above background. While the expression of the majority of the 28 human *ets* genes did not change over the 12-hr time course, we did observe changes in several family members. Upregulated genes included *etv5* (14-fold at 12 hr), *elf1* (2-fold at 8 hrs), and *ets2* (2-fold at 5 hr), while down-regulated genes included *ets1* (2-fold at 5 hr) and *etv6* (3.2-fold 5 hr) (Figure 3.8a). Western blot analysis tested protein levels of ETS1, ETS2, and ELF1 and indicated that protein levels were more dramatically affected than mRNA levels (Figure 3.8b). Western blots for ETV5 and ETV6 showed no change, thus, did not reflect the changed mRNA expression (data not shown), possibly due to post-translational regulation of these factors.

Increased ETS2 at 5 hr postactivation corresponds to increased expression of ETS1 specific gene targets at this time point. Thus we speculate that ETS2 might be a replacement for ETS1 at these sites. It was previously reported that there is a reciprocal relationship between ETS1 and ETS2 expression upon T cell activation [39,43]. Further experiments, which require development of more robust ETS2 antibodies, will explore this possibility.

Discussion

The work presented in this chapter supports a model of ETS factor redundancy wherein ETS factors co-expressed in a single cell share occupancy on an individual site and perform a redundant function. By perturbing the ETS

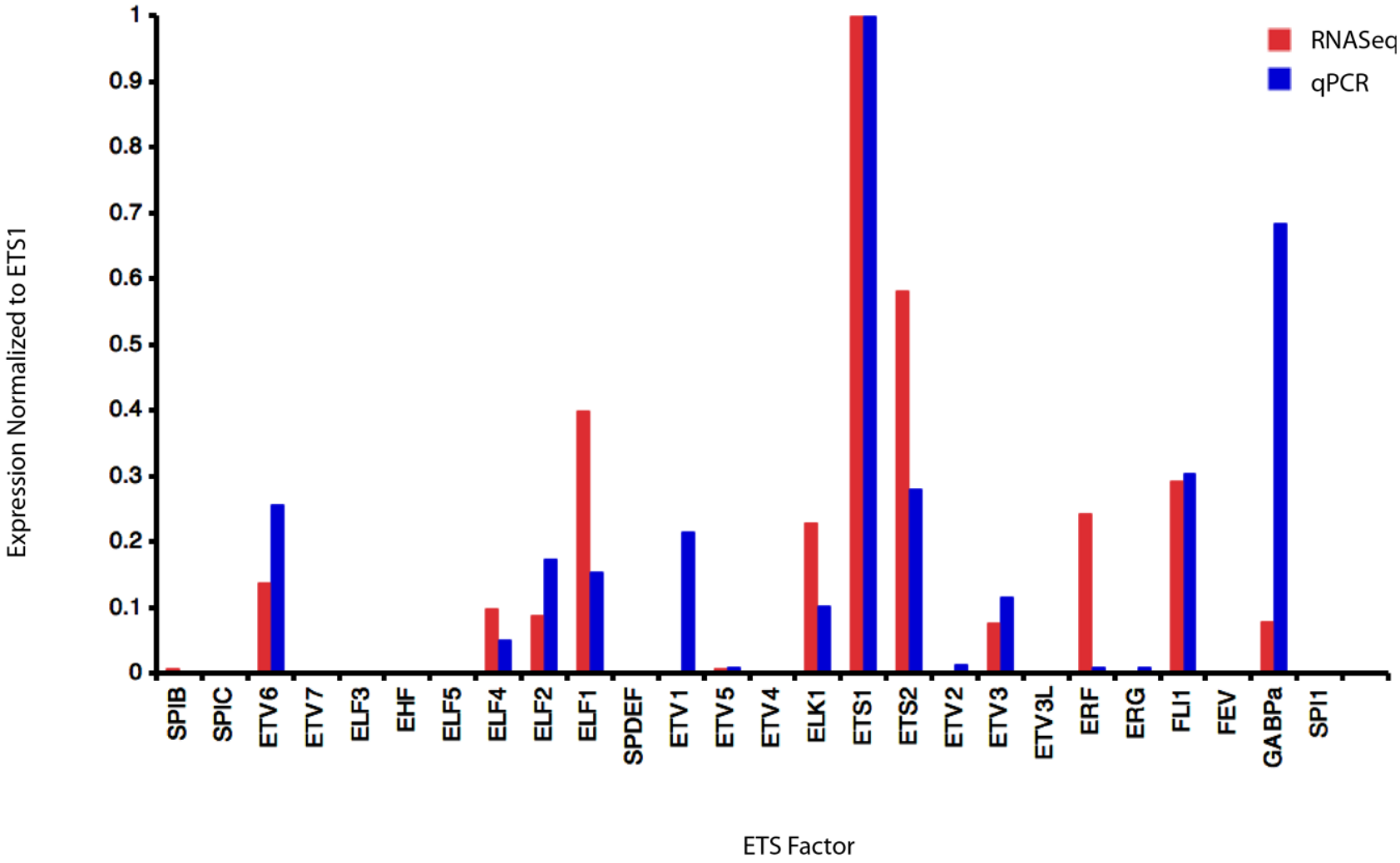


Figure 3.7 The ETS network in resting T cells Comparison of the resting state RNASeq data to mRNA copy numbers determined for each *ets* gene by RT-qPCR [14]. The two data sets were normalized to the most highly expressed ETS factors, in this case ETS1, and there was good correlation between the two methods.

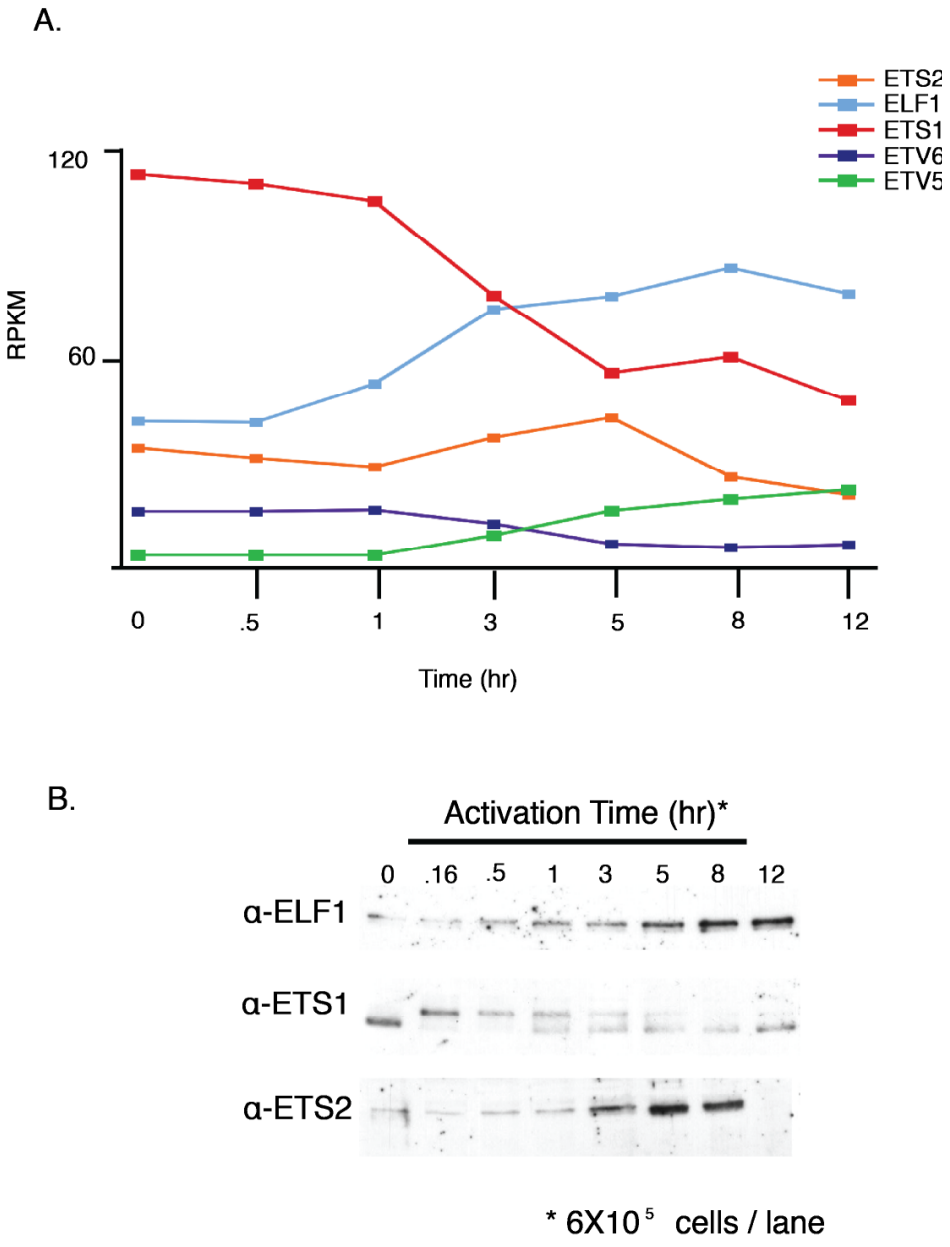


Figure 3.8 Perturbation of the ETS network (A) The expression of several ETS genes was affected by cell signaling. Upregulated genes included *etv5* (14-fold at 12 hr), *elf1* (2-fold at 8 hrs), and *ets2* (2-fold at 5 hr), and down-regulated genes included *ets1* (2-fold at 5 hr) and *etv6* (3.2-fold 5 hr). (B) Immunoblot analysis on whole cell extracts show that the increase in ETS1, ETS2, and ELF1 protein was more dramatic than the increase in mRNA. Immunoblot analysis for ETV5 and ETV6 showed no change in protein (data not shown).

network and profiling the expression of ETS target genes, distinct behavioral differences were identified for redundant ETS targets and specific ETS1 targets, further establishing that different recruitment motifs mediate distinct transcriptional functions of ETS1. Unexpectedly, ETS1 specific target genes were transcriptionally upregulated upon disruption of ETS1 binding, setting the stage for future experiments examining whether these sites are truly ETS1 specific. These findings add new complexity to the working model of ETS factor specificity and redundancy, and indicate that these delineations may not apply to all contexts. We put forth a model of ETS factor function which represents redundancy and specificity as a continuum rather than a dichotomy, particularly for closely-related family members (Figure 3.9).

The upregulation of ETS1-specific sites

The unexpected upregulation of ETS1 specific target genes could be mediated by a number of different mechanisms. While it is certainly possible that a non-ETS transcription factor mediates the observed upregulation, it is not unreasonable to predict regulation by a previously unconsidered ETS factor as the affected class of genes is correlated with an ETS binding motif, albeit a low-affinity motif.

Above, we have presented a straight forward model wherein the upregulation of ETS2 compensates for ETS1 disruption. Our lab previously found that ETS2 is minimally phosphorylated by CaMKII and that calcium signaling does not attenuate ETS2 DNA binding affinity (D. C. unpublished data). Upon T

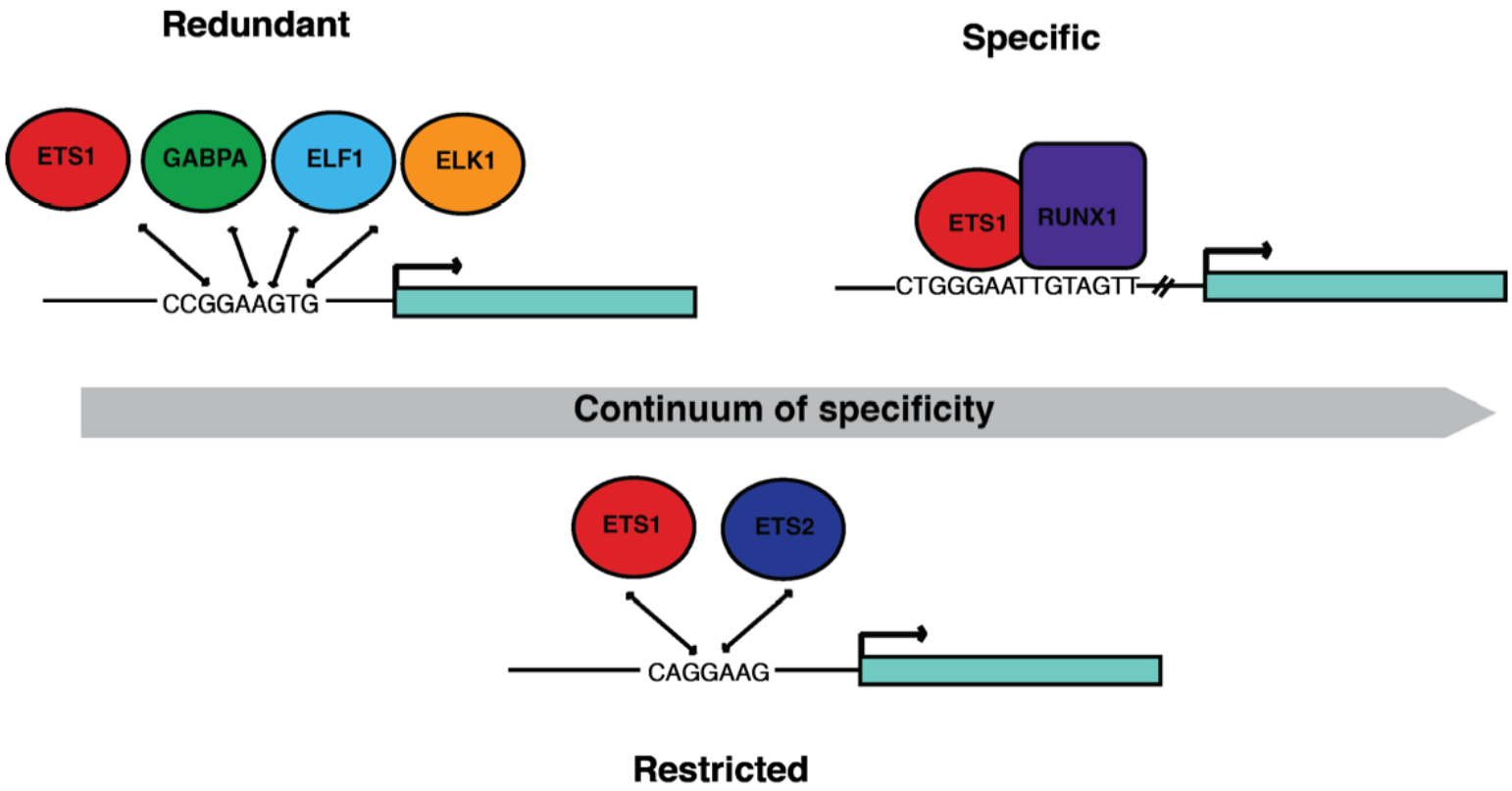


Figure 3.9 The ETS family functions on a continuum of specificity The work presented in this chapter suggests that the ETS family functions on a continuum of specificity. We propose a model of ETS factor function that includes redundant, restricted, and specific recruitment mechanisms. At high-affinity sites, ETS factors bind redundantly and regulate constitutive gene expression. At low-affinity sites, ETS factors may bind as a sub-set of ETS factors (restricted), or as an individual ETS factor (specific), and regulate inducible and cell type specific gene expression.

cell activation, the concentration of cyctoplasmic Ca^{2+} oscillates [44], which would disrupt transcription driven by ETS1 but presumably not ETS2. This difference between the two factors could facilitate ETS2 occupancy at these sites in the presence of calcium-dependant signaling, but it is unresolved whether the transcriptional properties of ETS2 are equivalent with those of ETS1.

Beyond a simple interchange of ETS2 for ETS1, RAS-MAPK signaling to one or both factors may play a role in the upregulation of this gene set. This chapter focused on calcium signaling to the ETS network, but T cell receptor engagement also activates the RAS-MAPK pathway [45]. The RAS-activated kinase ERK1/2 phosphorylates ETS1 and ETS2 in a conserved, unstructured region N-terminal to the PNT-domain [46]. This phosphorylation event increases the factors' transcriptional activity via an enhanced interaction between ETS1/2-PNT and the transcriptional co-activator CBP/p300, [47]. Our study found that ERK1/2 is no longer activated at the 5 hr time point (data not shown), but RAS signaling to ETS2, or residually bound ETS1, may still play a role in the observed upregulation of genes. We speculate that stably bound ETS1 may be modified by ERK-phosphorylation. Unfortunately, attempts to detect this phosphorylated ETS1 species at target sites with ChIP experiments and a phosphospecific antibody have not been successful.

Redundancy within the ETS family

Shifting the equilibrium of ETS1 occupancy affected the occupancy of other ETS factors, providing evidence that ETS factors share time on an

individual binding site. While expression data indicate that ETS factors perform a redundant function at HK promoters, the function is still unknown. *In vitro* functional studies have shown that ETS1 recruits the co-activator CBP/p300 to activate transcription [48]. A number of other ETS factors interact with CBP/p300, including ETS2 [49], PU.1 [50], GABPA [51], and ELK1 [52], and it has been suggested that recruitment of CBP/p300 to promote transcription could be the redundant function of ETS factors. Countering this possibility, the *in vivo* survey of CBP occupancy presented in Chapter 2 identified a CBP profile distinct from ETS1 and GABPA at HK promoters [25]. It is possible that ETS factors recruit CBP to HK promoters and subsequently 'hand-off' CBP to another molecule, such as acetylated histone via the CBP bromo-domain [53,54], but here we consider other mechanisms of redundancy.

Outside of the ETS domain, there is considerable diversity across the family. Perhaps the redundant function of the ETS family simply requires site occupancy. The strong occupancy of ETS factors at sites near a TSS could act as a place holder, maintaining or establishing open chromatin to facilitate transcriptional activation by other factors. Consistent with this possibility, we found that redundant ETS binding events occur in a nucleosome free region (NFR), about 25 bp upstream of the TSS [25]. Redundant preference for high affinity sites ensures a site is always occupied, and could serve to establish a transcriptionally permissive environment at genes requiring constitutive expression.

The promoters of HK genes are CpG rich and largely hypo-methylated (reviewed in [55]), which is essential for maintaining active chromatin and facilitating constitutive expression. The preferred ETS site contains a CpG dinucleotide in the core recognition site, and several ETS factors have been shown to be sensitive to DNA methylation [56,57]. Further, comparisons of methylation patterns and ETS factor occupancy reveal a correlation between the ETS site and hypo-methylation (data not shown). Thus, ETS factors could play a role in maintaining hypo-methylation at housekeeping promoters, contributing to the preservation of a transcriptionally permissive environment. This mechanism of protection by a transcription factor has been shown for the MLL, which protects a CpG-containing binding site in the Hoxa9 promoter from methylation and thereby promotes transcription [58].

Binding site sequence dictates behavior

We previously reported that unique recruitment motifs at ETS1 targets are correlated with distinct co-activator recruitment profiles, histone landscape, and genomic location [25]. Here, we have expanded on that finding to show that binding site sequence dictates distinct behaviors upon ETS network perturbation. Historically, transcription factor binding sites were viewed as platforms to recruit transcriptional machinery, not as tunable elements of transcriptional behavior. It is becoming clear, however, that most transcription factors bind to more than one distinct binding site, which may be associated with unique transcriptional mechanisms. Studies on the transcription factor glucocorticoid receptor (GR)

have shown that DNA can act as an allosteric-ligand and that site sequence can influence distinct conformations, and thereby transcriptional functions, of GR [59]. *In vitro* structural studies on PU.1 have shown that the DNA binding domain makes unique base-contacts with low-affinity and high-affinity sites [60], resulting in unique conformational complexes directed by different binding sites.

Considering the use of diverse DNA binding sites on a larger scale, the binding preferences of 104 transcription factors were assessed using protein-binding microarrays, revealing that approximately half of the factors considered bound to two or more distinct DNA motifs [61]. Our analysis of the behavior of unique ETS1 binding motifs has considered only two sites, but further genomic exploration of the ETS family will no doubt identify additional motifs and expand our current understanding of ETS factor function.

The continuum of ETS factor specificity

We report that previously identified 'ETS1 specific' binding events are associated with up regulated genes upon ETS1 disruption, leading us to hypothesize that ETS2, the ETS factor most related to ETS1, may be involved. The function of ETS1 and ETS2 is not entirely redundant, however, and mouse models predict a subset of unique functions for each factor [11,62,63,64]. *In vitro*, ETS2 was found to bind the consensus ETS site with 10-fold lower affinity than ETS1 (unpublished data), and the factors respond differentially to calcium signaling. Thus, further consideration of ETS1 and ETS2 can serve to address

the ways in which closely-related factors establish unique functions within the context of the family.

The work presented in this chapter highlights the importance of studying the ETS family as a network rather than as individual factors. Delineations of 'specific' and 'redundant' are no doubt too stringent for what is really a complex continuum of functionality within the family.

Materials and methods

T cell culturing and activation

Jurkat T cells (E6 line) were cultured in RPMI media supplemented with 10% FBS, 1X Pen/Strep, 10mM sodium pyruvate, and 10mM HEPEs buffer, and incubated at 37°C with 5% CO_2. T cell activation was recapitulated by treating confluent cultures with 2uM ionomycin (I9657Sigma) and 20uM PMA (Phorbol 12-myristate-13-acetate, P8139 Sigma) for the indicated incubation times.

Immunoblot analysis

Immunoblot analysis was performed by harvesting $10x10^6$ Jurkat cells, centrifuging at 500 x g for 5 min at 4°C, and resuspending whole cell pellets in SDS loading buffer (1uM DTT) to a concentration of 40,000 cells/uL. Resuspended pellets were sonicated and incubated at 95°C for 5 min. The extract of 600,000 cells was then loaded on a 10% polyacrylamide SDS gel and run at 200 V for 45 min at RT. Samples were transferred to a nitrocellulose membrane at 250 mA for 2 hr at 4°C. Membranes were blocked with 5% dry

milk:1X PBS (pH 7.4) for 30 min at RT. Primary antibody incubation was performed overnight at 4°C, and washed 4x for 5 min each (1X PBS, pH 7.4). Secondary antibody conjugations were performed for 1.5 hr at RT. Immunoblots were developed using SuperSignal® West Femto Maximum Sensitivity Substrate (Pierce, 34094). Antibodies used: ETS1 C20 (Santa Cruz; sc-350), ETS1T251P (Biosource; 44-1109G), ETS2 (Santa Cruz; sc-22803), ELF1 (Santa Cruz; sc-631), GABPa (Santa Cruz; sc-22810), and Tubulin (Santa Cruz; sc-5546).

ChIP experiments (qPCR)

ChIP experiments were performed as previously published [24]. In short, 50 x106 cells were cross linked with 1% formaldehyde at RT for 15 min. The crosslinking was quenched with 100mM Glycine for 5 min, then washed with 1X PBS (pH 7.4). The cells were then lysed, nuclei washed, and resuspended in a nuclei lysis buffer. The crosslinked chromatin was sonicated by a Misonix Ultrasonic processor using 10X 20 sec pulses at a Power Level of '4'. The sonicate was cleared for 10 min at >16,000 RPM, and the lysate of 10 $x10^6$ cells was used for each IP. Sonicate was added to 100 uL Dyna-beads (Invitrogen; Dynabeads® M-280) conjugated to ~5 ug antibody, and rotated at 4°C for 4 hr. IPs were washed 4X with 250mM salt and eluted in 10mM Tris (pH 7.9). The eluate was treated with RNaseA (Invitrogen; 12091-039) for 20 min at 37°C, and crosslinks were reversed overnight with ProteinaseK treatment (Invitrogen; AM2546). Samples were purified using a Qiaquick column (Qiagen; 28104), and the eluate was assayed for enrichment by qPCR using target specific primers

(Table 3.1) and the Roche480 system (Roche Applied Science; 04707516001) as previously described [24,25].

ChIPSeq analysis

ChIP DNA was size selected (100-500 bp fragments) and was single end sequenced with 36 cycles on an Illumina HiSeq GAII using standard protocols for preparing and sequencing libraries. ChIPSeq data was processed and analyzed as previously described [25] using the USeq software package (http://useq.sourceforge.net) [65]. Human genome sequence was obtained from the UCSC Genome Browser (March 2006, NCBI Build 36.1, HG18), and data was visualized with the Integrated Genome Browser (IGB, http://igb.bioviz.org/).

Expression profiling

RNA was isolated from 1X106 Jurkat cells using the Qiagen RNeasy Extraction kit (Qiagen; 74104) coupled with the QiaShredder Column (Qiagen; 79654) and an on-column DNase digestion (Qiagen; 79254). The isolated RNA was either assayed by RT-qPCR or by RNASeq. For RT-qPCR, RNA was reverse transcribed using SuperScript® III First-Strand Synthesis System and standard protocol (Invitrogen; 18080-051). The samples were RNaseH (Invitrogen; 18021-071) treated for 30 min at 37°C and cleaned using the Qiagen PCR cleanup kit. qPCR was performed using site specific primers (Table 3.2) and the Roche480 Light Cycler.

Table 3.1 ChIP qPCR Primers

Gene Target	Forward Primer	Reverse Primer
CDC2L2	CGCAGTTTCTTTTGGAGTCCTG	TCGGAACTCACCCCTACGGG
RPL32	GGCAAGCACTTCCGGCTCC	CCAAATTTCTACGTAGTCCCAAGG
RPS26	CAGCAGAAATGCTGAATGTAAAGG	CATGAGATCCCTACGCGGAC
GSK3A	GGAATAGTCCAGACCTGACC	CTAACCTTTCGTTCACACGCC
CD69	ATTTATTTTTAACCTGCAGGTGCC	TTGAAGAGGACAGAAGGTTGG
TCRB	GGAATCTAGGTATCCCAGATCC	CTGTGGCCTTTGATCTTGTGC
MMP9	AACCAGAACCACTTCCTGCC	TCCTGGCTCCCCAACCC

Table 3.2 RT- qPCR Primers

Gene Target	Forward Primer	Reverse Primer
CD82	GGGTCTGGTAGGTGGTGCC	GGGAGGGCTTTCCATTAAGG
CD83	CTCAGGTAACCAGACAGCC	CATTTTCCTTCCACCTCCTACC
GABPA	CACCATGCTGAATCAGAAGC	TGCTGAATTCCTTCATTACCC

For RNASeq, the extracted RNA was polyT selected and sequenced using the Illumina HiSeq GAII after a standard library preparation protocol. Sequences were aligned using the TopHat aligner (http://tophat.cbcb.umd.edu/) with a splice junction radius of 45 bp, and annotation was obtained from UCSC (March 2006, NCBI Build 36.1, HG18). Absolute expression levels were identified using the USeq bioinformatics package [65].

References

1. Nowick K, Stubbs L (2010) Lineage-specific transcription factors and the evolution of gene regulatory networks. Brief Funct Genomics 9: 65-78.

2. Guth SI, Wegner M (2008) Having it both ways: sox protein function between conservation and innovation. Cell Mol Life Sci 65: 3000-3018.

3. Suske G (1999) The Sp-family of transcription factors. Gene 238: 291-300.

4. Owen GI, Zelent A (2000) Origins and evolutionary diversification of the nuclear receptor superfamily. Cell Mol Life Sci 57: 809-827.

5. Rohs R, Jin X, West SM, Joshi R, Honig B, et al. (2010) Origins of specificity in protein-DNA recognition. Annu Rev Biochem 79: 233-269.

6. Hollenhorst PC, McIntosh LP, Graves BJ (2011) Genomic and biochemical insights into the specificity of ETS transcription factors. Annu Rev Biochem 80: 437-471.

7. Hsu T, Schulz RA (2000) Sequence and functional properties of Ets genes in the model organism Drosophila. Oncogene 19: 6409-6416.

8. Hart AH, Reventar R, Bernstein A (2000) Genetic analysis of ETS genes in C. elegans. Oncogene 19: 6400-6408.

9. Chen C, Ouyang W, Grigura V, Zhou Q, Carnes K, et al. (2005) ERM is required for transcriptional control of the spermatogonial stem cell niche. Nature 436: 1030-1034.

10. Laing MA, Coonrod S, Hinton BT, Downie JW, Tozer R, et al. (2000) Male sexual dysfunction in mice bearing targeted mutant alleles of the PEA3 ets gene. Mol Cell Biol 20: 9337-9345.

11. Muthusamy N, Barton K, Leiden JM (1995) Defective activation and survival of T cells lacking the Ets-1 transcription factor. Nature 377: 639-642.

12. Barton K, Muthusamy N, Fischer C, Ting CN, Walunas TL, et al. (1998) The Ets-1 transcription factor is required for the development of natural killer cells in mice. Immunity 9: 555-563.

13. Galang CK, Muller WJ, Foos G, Oshima RG, Hauser CA (2004) Changes in the expression of many Ets family transcription factors and of potential target genes in normal mammary tissue and tumors. J Biol Chem 279: 11281-11292.

14. Hollenhorst PC, Jones DA, Graves BJ (2004) Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. Nucleic Acids Res 32: 5693-5702.

15. Brown LA, Yang SH, Hair A, Galanis A, Sharrocks AD (1999) Molecular characterization of a zebrafish TCF ETS-domain transcription factor. Oncogene 18: 7985-7993.

16. John S, Marais R, Child R, Light Y, Leonard WJ (1996) Importance of low affinity Elf-1 sites in the regulation of lymphoid-specific inducible gene expression. J Exp Med 183: 743-750.

17. Mao X, Miesfeldt S, Yang H, Leiden JM, Thompson CB (1994) The FLI-1 and chimeric EWS-FLI-1 oncoproteins display similar DNA binding specificities. J Biol Chem 269: 18216-18222.

18. Ray-Gallet D, Mao C, Tavitian A, Moreau-Gachelin F (1995) DNA binding specificities of Spi-1/PU.1 and Spi-B transcription factors and identification of a Spi-1/Spi-B binding site in the c-fes/c-fps promoter. Oncogene 11: 303-313.

19. Shore P, Sharrocks AD (1995) The ETS domain transcription factors Elk-1 and SAP-1 exhibit differential DNA binding specificities. Nucleic Acids Res 23: 4698-4706.

20. Nye JA, Petersen JM, Gunther CV, Jonsen MD, Graves BJ (1992) Interaction of murine ets-1 with GGA-binding sites establishes the ETS domain as a new DNA-binding motif. Genes Dev 6: 975-990.

21. Wei GH, Badis G, Berger MF, Kivioja T, Palin K, et al. (2010) Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. EMBO J 29: 2147-2160.

22. Poon GMK, Macgregor RB (2003) Thermodynamic basis of sequence selectivity by the ETS domain of murine PU.1. Biophys J 84: 366A-366A.

23. Poon GMK, MacGregor RB (2004) A thermodynamic basis of DNA sequence selectivity by the ETS domain of murine PU.1. J Mol Biol 335: 113-127.

24. Hollenhorst PC, Shah AA, Hopkins C, Graves BJ (2007) Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. Genes Dev 21: 1882-1894.

25. Hollenhorst PC, Chandler KJ, Poulsen RL, Johnson WE, Speck NA, et al. (2009) DNA specificity determinants associate with distinct transcription factor functions. PLoS Genet 5: e1000778.

26. Boros J, O'Donnell A, Donaldson IJ, Kasza A, Zeef L, et al. (2009) Overlapping promoter targeting by Elk-1 and other divergent ETS-domain transcription factor family members. Nucleic Acids Res 37: 7368-7380.

27. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, et al. (2008) High-resolution mapping and characterization of open chromatin across the genome. Cell 132: 311-322.

28. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, et al. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129: 823-837.

29. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, et al. (2005) A high-resolution map of active promoters in the human genome. Nature 436: 876-880.

30. Kim TH, Barrera LO, Qu C, Van Calcar S, Trinklein ND, et al. (2005) Direct isolation and identification of promoters in the human genome. Genome Res 15: 830-839.

31. Rabault B, Ghysdael J (1994) Calcium-induced phosphorylation of ETS1 inhibits its specific DNA binding activity. J Biol Chem 269: 28143-28151.

32. Cowley DO, Graves BJ (2000) Phosphorylation represses Ets-1 DNA binding by reinforcing autoinhibition. Genes Dev 14: 366-376.

33. Pufall MA, Lee GM, Nelson ML, Kang HS, Velyvis A, et al. (2005) Variable control of Ets-1 DNA binding by multiple phosphates in an unstructured region. Science 309: 142-145.

34. Lee GM, Pufall MA, Meeker CA, Kang HS, Graves BJ, et al. (2008) The affinity of Ets-1 for DNA is modulated by phosphorylation through transient interactions of an unstructured region. J Mol Biol 382: 1014-1030.

35. Liu H, Grundstrom T (2002) Calcium regulation of GM-CSF by calmodulin-dependent kinase II phosphorylation of Ets1. Mol Biol Cell 13: 4497-4507.

36. Grenningloh R, Miaw SC, Moisan J, Graves BJ, Ho IC (2008) Role of Ets-1 phosphorylation in the effector function of Th cells. Eur J Immunol 38: 1700-1705.

37. Koizumi S, Fisher RJ, Fujiwara S, Jorcyk C, Bhat NK, et al. (1990) Isoforms of the human ets-1 protein: generation by alternative splicing and differential phosphorylation. Oncogene 5: 675-681.

38. Jorcyk CL, Watson DK, Mavrothalassitis GJ, Papas TS (1991) The human ETS1 gene: genomic structure, promoter characterization and alternative splicing. Oncogene 6: 523-532.

39. Bhat NK, Thompson CB, Lindsten T, June CH, Fujiwara S, et al. (1990) Reciprocal expression of human *ETS1* and *ETS2* genes during T-cell activation: regulatory role for the protooncogene *ETS1*. Proc Natl Acad Sci USA 87: 3723-3727.

40. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5: 621-628.

41. Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, et al. (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods 5: 829-834.

42. Ghisletti S, Barozzi I, Mietton F, Polletti S, De Santa F, et al. (2010) Identification and characterization of enhancers controlling the inflammatory gene expression program in macrophages. Immunity 32: 317-328.

43. Fujiwara S, Fisher RJ, Bhat NK, de la Espina SMD, Papas TS (1988) A short-lived nuclear phosphoprotein encoded by the human *ets*-2 proto-oncogene is stabilized by activation of protein kinase C. Mol Cell Biol 8: 4700-4706.

44. Guse AH, da Silva CP, Berg I, Skapenko AL, Weber K, et al. (1999) Regulation of calcium signalling in T lymphocytes by the second messenger cyclic ADP-ribose. Nature 398: 70-73.

45. Bar-Sagi D, Hall A (2000) Ras and Rho GTPases: a family reunion. Cell 103: 227-238.

46. Yang B-S, Hauser CA, Henkel G, Colman MS, Van Beveren C, et al. (1996) Ras-mediated phosphorylation of a conserved threonine residue enhances the transactivation activities of c-Ets-1 and c-Ets-2. Mol Cell Biol 16: 538-547.

47. Foulds CE, Nelson ML, Blaszczak AG, Graves BJ (2004) Ras/mitogen-activated protein kinase signaling activates Ets-1 and Ets-2 by CBP/p300 recruitment. Mol Cell Biol 24: 10954-10964.

48. Yang C, Shapiro LH, Rivera M, Kumar A, Brindle PK (1998) A role for CREB binding protein and p300 transcriptional coactivators in Ets-1 transactivation functions. Mol Cell Biol 18: 2218-2229.

49. Jayaraman G, Srinivas R, Duggan C, Ferreira E, Swaminathan S, et al. (1999) p300/cAMP-responsive element-binding protein interactions with ets-1 and ets-2 in the transcriptional activation of the human stromelysin promoter. J Biol Chem 274: 17342-17352.

50. Yamamoto H, Oikawa T (1999) Interaction of transcription factor PU. 1 with coactivator CBP. Tanpakushitsu Kakusan Koso 44: 1389-1395.

51. Kang HS, Nelson ML, Mackereth CD, Scharpf M, Graves BJ, et al. (2008) Identification and structural characterization of a CBP/p300-binding domain from the ETS family transcription factor GABP alpha. J Mol Biol 377: 636-646.

52. Janknecht R, Ernst WH, Pingoud V, Nordheim A (1993) Activation of ternary complex factor Elk-1 by MAP kinases. Embo J 12: 5097-5104.

53. Mujtaba S, He Y, Zeng L, Yan S, Plotnikova O, et al. (2004) Structural mechanism of the bromodomain of the coactivator CBP in p53 transcriptional activation. Mol Cell 13: 251-263.

54. Mujtaba S, Zeng L, Zhou MM (2007) Structure and acetyl-lysine recognition of the bromodomain. Oncogene 26: 5521-5527.

55. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25: 1010-1022.

56. Yokomori N, Kobayashi R, Moore R, Sueyoshi T, Negishi M (1995) A DNA methylation site in the male-specific P450 (Cyp 2d-9) promoter and binding of the heteromeric transcription factor GABP. Mol Cell Biol 15: 5355-5362.

57. Lucas ME, Crider KS, Powell DR, Kapoor-Vazirani P, Vertino PM (2009) Methylation-sensitive regulation of TMS1/ASC by the Ets factor, GA-binding protein-alpha. J Biol Chem 284: 14698-14709.

58. Erfurth FE, Popovic R, Grembecka J, Cierpicki T, Theisler C, et al. (2008) MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression. Proc Natl Acad Sci U S A 105: 7517-7522.

59. Meijsing SH, Pufall MA, So AY, Bates DL, Chen L, et al. (2009) DNA binding site sequence directs glucocorticoid receptor structure and activity. Science 324: 407-410.

60. Poon GM (2012) Sequence discrimination by the DNA-binding domain of the ETS-family transcription factor PU.1 is linked to specific hydration of the protein-DNA interface. J Biol Chem.

61. Badis G, Berger MF, Philippakis AA, Talukder S, Gehrke AR, et al. (2009) Diversity and complexity in DNA recognition by transcription factors. Science 324: 1720-1723.

62. Bories JC, Willerford DM, Grevin D, Davidson L, Camus A, et al. (1995) Increased T-cell apoptosis and terminal B-cell differentiation induced by inactivation of the Ets-1 proto-oncogene. Nature 377: 635-638.

63. Barton K, Muthusamy N, Fischer C, Ting C-N, Walunas T, et al. (1998) The Ets-1 transcription factor is required for the development of natural killer cells in mice. Immunity 9: 555-563.

64. Yamamoto H, Flannery ML, Kupriyanov S, Pearce J, McKercher SR, et al. (1998) Defective trophoblast function in mice with a targeted mutation of Ets2. Genes Dev 12: 1315-1326.

65. Nix DA, Courdy SJ, Boucher KM (2008) Empirical methods for controlling false positives and estimating confidence in ChIP-Seq peaks. BMC Bioinformatics 9: 523.

CHAPTER 4

SUMMARY, DISCUSSION, AND FUTURE DIRECTIONS

A genomic approach to the ETS network

At all levels of biological complexity – from a single cell to high order organisms – transcriptional regulation dictates the complex and coordinated gene expression programs required for proper growth, development, and function. Much of this diversity is achieved through the use of distinct transcription factors (TFs). Most TFs exist in the context of highly related families that exhibit functional specificity in spite of a high degree of homology. Understanding how related factors establish unique functions is central to the understanding of transcriptional regulation and cell biology as a whole.

Summary of genome-wide ETS studies

Genome wide studies have begun to reveal the global distribution of ETS factor binding events, and these studies have shown that ETS factors are recruited to DNA by both redundant and specific regulatory elements. With a focus on ETS1, the work presented in this dissertation refines the model of ETS family function by identifying architectural, genomic, biological, and behavioral differences between different types of ETS1 occupied regulatory elements.

In Chapter 2, we established that ETS-redundant and ETS1-specific regulatory elements are associated with distinct chromatin landscapes, DNA recruitment motifs, and ontological gene sets. Using a genome wide approach, we expanded our understanding of specific ETS1-sites to include a large number previously unidentified distal transcriptional enhancers. These ETS1-specific binding events were found to be mediated by a degenerate ETS binding motif,

and in many cases the nearest gene was associated with T cell biology. Overlaying histone mark profiles on ETS-factor occupancy data, we found that ETS-redundant and ETS1-specific regulatory elements are correlated with marks of active promoters and transcriptional enhancers, respectively. We also found a striking co-localization of ETS1 and the transcriptional co-activator CBP at enhancers, but not at promoters, suggesting that ETS1 may behave differently at these two recruitment sequences.

In Chapter 3, we disrupted ETS1 binding *in vivo* to reveal distinct behaviors for ETS-redundant and ETS1-specific target genes, illustrating the distinct biological functions mediated by ETS1. At redundant promoters, disruption of ETS1 occupancy revealed the first example of dynamic time sharing by ETS1 and GABPA, establishing the basis of redundancy within the ETS family. Further, this time sharing had no effect on the expression of nearby genes, suggesting functional redundancy for ETS factors at these sites. Time sharing was not observed at previously identified ETS1-specific sites, indicating that specificity is not an artifact of competition. Unexpectedly, the average transcription of genes near ETS1-specific binding events was increased upon ETS1 disruption, and we hypothesize that this could be due to the concurrent increase in ETS2 protein. These results have important implications for understanding how ETS1 functionally mediates different transcriptional programs, and suggests a model of ETS family function wherein specificity is a continuum rather than a dichotomy.

Discussion

Refining the model of ETS family function

The work presented in this dissertation uses data sets generated by the massively parallel DNA sequencing techniques ChIPSeq and RNASeq to refine the model of ETS family function *in vivo*. At the commencement of this work, ChIP-on-chip using promoter microarrays had revealed a subset of redundant and specific ETS factor binding events in Jurkat T cells [1]. While these studies provided a conceptual framework for understanding and describing ETS factor binding in the cell, the use of probe-based techniques like microarray limited both the resolution and comprehensiveness of the working model. As presented in Chapter 2, the use of ChIPSeq – a massively parallel DNA sequencing technique which does not rely on the use of predefined probes – allowed us to interrogates factor occupancy genome-wide rather than at a subset of preselected genomic loci. We expanded our understanding of the ETS family model to include a large number of distal transcriptional enhancers, including 12,283 ETS1-bound enhancers (>500 bp from TSS) [2] which would not have been identified by promoter microarrays.

Transcription factor binding does not presage transcription factor function, however, and functional validation of ChIPSeq data is required. Correlating *in vivo* factor occupancy with endogenous gene expression provides a biologically relevant way to confirm functionality and has identified novel functional mechanisms for transcription factors like estrogen receptor (ER) and p53 [3,4].

While it was clear that determining the expression profiles of putative ETS targets would add detail to the model of ETS factor function, the probe-based expression microarrays available at the commencement of this work presented a major obstacle for these studies. Inherent differences in microarray probe efficiency require the comparison of experimental data sets to null or control sets, which limits the utility of microarry to asses the expression of widely expressed genes like housekeeping genes. Other methods of establishing transcription, such as correlation of HK gene promoters with DNaseI hypersensitivity, H3K4me3, and PolII occupancy, had suggested that genes near redundant ETS sites were actively expressed [5,6,7,8], but direct evidence of active transcription was still lacking. In Chapter 3, we presented evidence that both redundant and specific ETS factor binding is correlated with positive transcription of neighboring genes. Using RNASeq – a quantitative, rather than qualitative, method of examining global gene expression – we established that 78% of genes near redundant ETS binding events and 66% of genes near ETS1 specific binding events are actively transcribed. This finding supports our prediction that ETS factors function as transcriptional activators at these sites and refines the model of the ETS family.

Taken together, genome wide sequencing techniques have added new features to the portrait of ETS family function. The overlay of new data sets for additional TFs, DNA methylation, and nucleosome positioning. Below, we discuss several of the outstanding questions related to the working model of the ETS family, including the functional mechanisms of redundancy and specificity, and

the implications of our finding in Chapter 3 that ETS1 and ETS2 may function on a continuum of specificity.

The redundant role of the ETS family

The degree of redundancy between ETS1, GABPA, and ELF1 was surprising, but previous to our finding there was evidence that ETS factors play a role in HK gene expression. HK gene promoters are CpG rich and an 8-mer high affinity ETS site, 5'-VCCGGAARY-3', is over represented in HK promoters [9,10]. Further, CpG rich promoters containing the 5'-VCCGGAARY-3' site frequently have high PolII occupancy across multiple cell types [7,8], suggesting constitutive expression of nearby genes. These observations independently validate our finding that high affinity ETS factor binding *in vivo* is correlated with actively transcribed HK genes.

In addition to establishing the transcription of genes near redundant binding sites, we have shown that expression levels are unchanged as ETS1 occupancy is disrupted and GABPA occupancy is increased. This finding suggests functional redundancy at ETS consensus sites, but it is unclear if the redundancy of ETS factors is entirely universal across the family. To date, genome wide studies of ETS factors have surveyed the occupancy of ETS factors with relatively divergent ETS domains, and have consistently found striking redundancy at HK promoters [2,11,12,13]. These findings suggest a universally redundant function of ETS factors, but systemic assessment of the

family must be performed to directly establish redundancy for all ETS family members.

Perhaps more significantly, however, the mechanism of redundant function is not understood, and there are many possible roles for ETS factors at HK promoters. Co-activator recruitment is an attractive possibility for the redundant function of ETS factors as many family members, including ETS1 [14], ETS2 [15], PU.1 [16], GABPA [17], and ELK1 [18], interact with the transcriptional co-activator CBP/p300. However, many of these ETS factors interact with CBP/p300 via the PNT interaction domain, and it is unlikely that all ETS factors can form interactions with CBP/p300. Further, fine mapping of our ChIPSeq data found that ETS1 and GABPA do not co-localize with CBP at HK promoters [2]. It is certainly possible that ETS factors recruit CBP to HK promoters but do not remain physically associated after recruitment, but other mechanisms of redundancy should be considered.

The most conserved structural domain across the ETS family is the ETS domain, and it is possible that redundancy is mediated through this domain. This could be established through a conserved protein partnership mediated by the ETS domain, and while pull-down experiments using the ETS domains of various ETS factors have yet to observe evidence of a common interacting protein, it remains an interesting possibility.

It is also possible that the ETS domain could mediate redundant function simply through the act of physically occupying a DNA binding site and protecting

it from methylation. The promoters of HK genes are CpG rich and largely hypo-methylated (reviewed in [19]), which is essential for maintaining active chromatin and facilitating constitutive expression. Several ETS factors have been shown to be sensitive to CpG methylation [20,21], and comparisons of methylation patterns and ETS factor occupancy reveal a correlation between the consensus ETS site and hypo-methylation (data not shown). Thus, ETS factors could play a role in maintaining hypo-methylation at housekeeping promoters and establishing a transcriptionally permissive environment. This mechanism of protection by a transcription factor has been shown for MLL, which protects a CpG-containing binding site in the Hoxa9 promoter from methylation and thereby promotes transcription [22].

Yet another possible mechanism of ETS factor function at redundant promoters could be through the positioning of promoter nucleosomes. If ETS factors can compete with nucleosome occupancy, they may play a role in establishing the nucleosome free region (NFR) at HK promoters. Bioinformatic analysis has revealed that redundant ETS binding events occur in the NFR, about 25 bp upstream of the TSS [2]. By helping to establish open chromatin at the TSS, ETS factors may promote binding of other TFs to their binding sites [23]. Future experiments will address the possibility of ETS factor competition with nucleosomes through bioinformatic and *in vitro* approaches (see below).

Specific functions of the ETS family

The potential for competition between ETS factors and nucleosomes may also be relevant to ETS1 function at transcriptional enhancers. Bioinformatic analysis reveals that many TF binding sites occur between nucleosomes [24], and in Chapter 2 we reported that well positioned nucleosomes flank ETS1 binding at enhancer regions. We hypothesize that the role of ETS1 at these sites may be to 'set-up' the transcriptional enhancer for activation, positioning nucleosomes such that factors involved in T cell activation can immediately access binding sites upon signaling. Indeed, it is thought that inducible recruitment of NF-kB is guided by the pre-set chromatin state [25], and ETS1 could play a role in determining the resting state nucleosome framework. It is also possible that the clearing of ETS1 from transcriptional enhancers upon activation may be a key step in the regulation of these gene targets.

Alternatively, ETS1 could set up T cell specific enhancers by recruiting additional factors to the site, establishing a transcriptionally permissive environment. In Chapter 2, fine mapping of enhancer sites revealed co-localization of ETS1 with CBP and the TF RUNX1, suggesting the establishment of an 'enhancesome' complex at these sites. ETS1 or the associated factors could play a role in the recruitment of histone modifying enzymes to establish the observed pattern of H3K4me1/3 at transcriptional enhancers [6,26].

The observation that the majority of ETS1 specific occupancy occurs at distal enhancers was significant to the understanding of ETS family function, and

is consistent with meta-analysis of genomic data sets indicating that distal regulatory elements mediate cell type specificity [27]. Interestingly, the most over represented binding site at ETS1 specific binding events is 5'-CAGGAA/T-3', which unlike the ETS redundant site, does not contain a CpG dinucleotide [1]. It has been suggested that in general, transcription factor binding sites which contain a CpG are constitutively expressed while binding sites which do not contain a CpG are involved in regulated, and often inducible, gene expression [10]. This suggestion is consistent with our findings for the ETS family, and consideration of non-CpG containing ETS sites may contribute to the identification of other ETS factor specific DNA recruitment motifs.

Specificity and redundancy within ETS subgroups- the ETS family continuum

In Chapter 3, we reported that previously identified 'ETS1 specific' binding events are associated with upregulated genes upon ETS1 disruption, leading us to hypothesize that ETS2, the ETS factor most related to ETS1, may play a role. ETS1 and ETS2 are not entirely interchangeable, however, and mouse models predict a subset of unique functions for each factor [28,29,30,31]. *In vitro*, ETS2 was found to bind the ETS consensus site with 10-fold lower affinity than ETS1, and the factors respond differentially to calcium signaling (D.O. Cowley, PhD thesis). Thus, further consideration of ETS1 and ETS2 can serve as a model-within-a-model, addressing how closely-related factors establish redundant and specific functions within the greater context of the family.

There are other examples of closely-related ETS factors being implicated in redundant but cell-type specific roles. In Ewing's sarcoma, the closely related ETS factors ERG, FLI, and FEV have been implicated in oncogenesis, but so has the more distantly related ETV1 [32,33,34,35]. Within the PEA3 subfamily, ETV1, ETV4, and ETV5 have been implicated in a redundant oncogenic function in prostate cells [36], but again a more distantly related ETS factor, ERG, has been implicated in this role. These oncogenic ETS represent two disparate clades in the family. These findings suggest that while ETS factors function on a continuum of specificity, the distribution of the continuum may not be solely influenced by evolutionary conservation.

These examples highlight the importance of studying the ETS family as a network rather than as individual factors, and the work presented in this thesis has proven the utility of studying the ETS family in a dynamic system. Delineations of 'specific' and 'redundant' are no doubt too stringent for what is really a complex continuum of functionality within the family, and future work in the T cell and other systems will no doubt reveal additional subtleties to the model of ETS factor function *in vivo*.

Future directions

The work presented in this dissertation frames a number of interesting future directions. Immediate future work will address the possible interchange of ETS1 and ETS2 at a subset of genomic loci. The time course expression data could also be examined for further expression patterns correlated with

established or putative ETS binding events. Additionally, addressing the interplay between ETS factor binding and nucleosome occupancy may provide mechanistic insight into the role of ETS factors in genome organization, both at promoters and enhancers.

Determine the *in vitro* binding affinity of ETS2
at enhancer sites

The hypothesis presented at the end of Chapter 3 suggests that ETS1 and ETS2 may bind redundantly to the degenerate ETS site found at many enhancer regions. The viability of this hypothesis can be tested by determining the *in vitro* DNA binding affinity of ETS2 for the enhancer site. We have found that ETS2 binds to the consensus ETS site with 10-fold lower affinity than ETS1 (D.O. Cowley, PhD thesis), and lower affinity for an already weak site may prevent significant ETS2 binding. Thus, establishing the feasibility of this hypothesis will require an examination of ETS2 binding affinity for the enhancer ETS site.

Determine the *in vivo* genomic occupancy of ETS2
at enhancer sites

Another approach to testing the hypothesis that ETS2 compensates for ETS1 at a subset of low-affinity ETS sites is to determine *in vivo* occupancy by ChIP qPCR. Alternatively, a genomic approach could be taken to this experiment and ETS2 ChIPSeq could be performed on activated cells. The advantage of the latter approach is the potential to discover ETS2 binding events unique from the

previously identified ETS1 binding events. A caveat to this experiment is that commercially available ETS2 antibodies have not been successful for ChIP.

Identify inducible genes near non-ETS1 occupied ETS binding sites

The work presented in this dissertation has focused largely on the specific functions of ETS1 in the dynamic T cell system. In Chapter 3, we presented data that the expression of ETS factors other than ETS1 and ETS2 are affected upon system perturbation. Notably, ETV5 is upregulated ~20-fold while ETV6, which is thought to be a transcriptional repressor, is downregulated. By identifying a set of changed genes with nearby ETS binding sites – though not necessarily ETS1 occupancy – new levels of dynamic regulation in the ETS family could be uncovered.

Perform *in vitro* competition assays between ETS factors and nucleosomes

We have suggested that the mechanism of redundancy for the ETS family is to permanently occupy high affinity sites and effectively occlude nucleosomes. We could test this hypothesis by constructing an *in vitro* chromatin template with a nucleosome-buried high-affinity ETS site. This chromatinized template could then be used in competition experiments with a minimal-ETS domain, revealing kinetic data for competition between ETS factors and nucleosomes. It poses a

caveat, however, in that chromatin templates *in vitro* may not represent the biological placement of chromatin *in vivo*.

Bioinformatically assess the model of ETS factor- nucleosome competition

Evidence for this mechanism could also be determined bioformatically. If ETS factors bind high-affinity sites to position nucleosomes, we predict that high-affinity ETS sites should occur either between nucleosomes, where an ETS factor has already competed away nearby nucleosomes, or buried within a nucleosome, where the site is inaccessible. Binding sites located at nucleosome-borders would be partially accessible due to nucleosome dynamics, or chromatin breathing, and binding of ETS factors would drive these sites to become inter-nucleosomal. Thus, if we observe high-affinity ETS sites at nucleosome-borders, the 'competition' mechanism of redundancy is less likely.

Final summary

In summary, the studies described within this dissertation identified distinct characteristics for redundant and specific ETS1 regulatory elements. Using massively parallel DNA sequencing techniques, this work adds new levels of detail to the model of ETS family function. In Chapter 2, this thesis presented the first genome wide occupancy profile of ETS1 and CBP, and comparison of this data to other genome wide data sets highlighted the distinct characteristics of redundant promoters and specific enhancers. The work presented in Chapter 3 of

this thesis revealed the first example of dynamic sharing of ETS factors *in vivo*, and suggested a molecular basis for the widely observed redundancy across the family. Expression data identified distinct behaviors for genes near different binding sites, further indicating that biological function is correlated with DNA binding site sequence. These results serve as a framework for future studies investigating the specificity and redundancy of ETS factors, particularly closely-related ETS factors, in an *in vivo* context.

References

1. Hollenhorst PC, Shah AA, Hopkins C, Graves BJ (2007) Genome-wide analyses reveal properties of redundant and specific promoter occupancy within the ETS gene family. Genes Dev 21: 1882-1894.

2. Hollenhorst PC, Chandler KJ, Poulsen RL, Johnson WE, Speck NA, et al. (2009) DNA specificity determinants associate with distinct transcription factor functions. PLoS Genet 5: e1000778.

3. Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, et al. (2005) Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 122: 33-43.

4. Wei CL, Wu Q, Vega VB, Chiu KP, Ng P, et al. (2006) A global map of p53 transcription-factor binding sites in the human genome. Cell 124: 207-219.

5. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, et al. (2008) High-resolution mapping and characterization of open chromatin across the genome. Cell 132: 311-322.

6. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE, et al. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129: 823-837.

7. Kim TH, Barrera LO, Zheng M, Qu C, Singer MA, et al. (2005) A high-resolution map of active promoters in the human genome. Nature 436: 876-880.

8. Kim TH, Barrera LO, Qu C, Van Calcar S, Trinklein ND, et al. (2005) Direct isolation and identification of promoters in the human genome. Genome Res 15: 830-839.

9. FitzGerald PC, Shlyakhtenko A, Mir AA, Vinson C (2004) Clustering of DNA sequences in human promoters. Genome Res 14: 1562-1574.

10. Rozenberg JM, Shlyakhtenko A, Glass K, Rishi V, Myakishev MV, et al. (2008) All and only CpG containing sequences are enriched in promoters abundantly bound by RNA polymerase II in multiple tissues. BMC Genomics 9: 67.

11. Hollenhorst PC, Jones DA, Graves BJ (2004) Expression profiles frame the promoter specificity dilemma of the ETS family of transcription factors. Nucleic Acids Res 32: 5693-5702.

12. Boros J, O'Donnell A, Donaldson IJ, Kasza A, Zeef L, et al. (2009) Overlapping promoter targeting by Elk-1 and other divergent ETS-domain transcription factor family members. Nucleic Acids Res 37: 7368-7380.

13. Wei GH, Badis G, Berger MF, Kivioja T, Palin K, et al. (2010) Genome-wide analysis of ETS-family DNA-binding in vitro and in vivo. EMBO J 29: 2147-2160.

14. Yang SH, Yates PR, Whitmarsh AJ, Davis RJ, Sharrocks AD (1998) The Elk-1 ETS-domain transcription factor contains a mitogen-activated protein kinase targeting motif. Mol Cell Biol 18: 710-720.

15. Jayaraman G, Srinivas R, Duggan C, Ferreira E, Swaminathan S, et al. (1999) p300/cAMP-responsive element-binding protein interactions with ets-1 and ets-2 in the transcriptional activation of the human stromelysin promoter. J Biol Chem 274: 17342-17352.

16. Yamamoto H, Oikawa T (1999) Interaction of transcription factor PU. 1 with coactivator CBP. Tanpakushitsu Kakusan Koso 44: 1389-1395.

17. Kang HS, Nelson ML, Mackereth CD, Scharpf M, Graves BJ, et al. (2008) Identification and structural characterization of a CBP/p300-binding domain from the ETS family transcription factor GABP alpha. J Mol Biol 377: 636-646.

18. Janknecht R, Ernst WH, Pingoud V, Nordheim A (1993) Activation of ternary complex factor Elk-1 by MAP kinases. Embo J 12: 5097-5104.

19. Deaton AM, Bird A (2011) CpG islands and the regulation of transcription. Genes Dev 25: 1010-1022.

20. Yokomori N, Kobayashi R, Moore R, Sueyoshi T, Negishi M (1995) A DNA methylation site in the male-specific P450 (Cyp 2d-9) promoter and binding of the heteromeric transcription factor GABP. Mol Cell Biol 15: 5355-5362.

21. Lucas ME, Crider KS, Powell DR, Kapoor-Vazirani P, Vertino PM (2009) Methylation-sensitive regulation of TMS1/ASC by the Ets factor, GA-binding protein-alpha. J Biol Chem 284: 14698-14709.

22. Erfurth FE, Popovic R, Grembecka J, Cierpicki T, Theisler C, et al. (2008) MLL protects CpG clusters from methylation within the Hoxa9 gene, maintaining transcript expression. Proc Natl Acad Sci U S A 105: 7517-7522.

23. Mirny LA (2010) Nucleosome-mediated cooperativity between transcription factors. Proc Natl Acad Sci U S A 107: 22534-22539.

24. Schones DE, Cui K, Cuddapah S, Roh TY, Barski A, et al. (2008) Dynamic regulation of nucleosome positioning in the human genome. Cell 132: 887-898.

25. Smale ST (2010) Selective transcription in response to an inflammatory stimulus. Cell 140: 833-844.

26. Heintzman ND, Hon GC, Hawkins RD, Kheradpour P, Stark A, et al. (2009) Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature 459: 108-112.

27. Visel A, Blow MJ, Li Z, Zhang T, Akiyama JA, et al. (2009) ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature 457: 854-858.

28. Bories JC, Willerford DM, Grevin D, Davidson L, Camus A, et al. (1995) Increased T-cell apoptosis and terminal B-cell differentiation induced by inactivation of the Ets-1 proto-oncogene. Nature 377: 635-638.

29. Muthusamy N, Barton K, Leiden JM (1995) Defective activation and survival of T cells lacking the Ets-1 transcription factor. Nature 377: 639-642.

30. Barton K, Muthusamy N, Fischer C, Ting C-N, Walunas T, et al. (1998) The Ets-1 transcription factor is required for the development of natural killer cells in mice. Immunity 9: 555-563.

31. Yamamoto H, Flannery ML, Kupriyanov S, Pearce J, McKercher SR, et al. (1998) Defective trophoblast function in mice with a targeted mutation of Ets2. Genes Dev 12: 1315-1326.

32. Giovannini M, Biegel JA, Serra M, Wang JY, Wei YH, et al. (1994) EWS-erg and EWS-Fli1 fusion transcripts in Ewing's sarcoma and primitive neuroectodermal tumors with variant translocations. J Clin Invest 94: 489-496.

33. Mackintosh C, Madoz-Gurpide J, Ordonez JL, Osuna D, Herrero-Martin D (2010) The molecular pathogenesis of Ewing's sarcoma. Cancer Biol Ther 9: 655-667.

34. Jeon IS, Davis JN, Braun BS, Sublett JE, Roussel MF, et al. (1995) A variant Ewing's sarcoma translocation (7;22) fuses the EWS gene to the ETS gene ETV1. Oncogene 10: 1229-1234.

35. Ng TL, O'Sullivan MJ, Pallen CJ, Hayes M, Clarkson PW, et al. (2007) Ewing sarcoma with novel translocation t(2;16) producing an in-frame fusion of FUS and FEV. J Mol Diagn 9: 459-463.

36. Hollenhorst PC, Ferris MW, Hull MA, Chae H, Kim S, et al. (2011) Oncogenic ETS proteins mimic activated RAS/MAPK signaling in prostate cells. Genes Dev 25: 2147-2157.